DOCUMENT RESUME

ED 033 628 FL 001 448

AUTHOR	Torrey, Jane W.
TITLE	The Learning of Grammar: An Experimental Study. Experimental Studies on the Learning of Language, Progress Report II, Revised.
INSTITUTION	Connecticut Coll., New London. Dept. of Psychology.
Spons Agency	National Inst. of Mental Health (DHEW), Bethesda, Md.; Public Health Service (DHEW), Washington, D.C.
Pub Date	Jan 66
Note	116p.
EDRS Price	EDRS Price MF-$0.50 HC-$5.90
Descriptors	Behavioral Objectives, Behavioral Science Research, *Educational Experiments, *Grammar, Grammar Translation Method, Language Instruction, Language Research, Linguistics, *Pattern Drills (Language), Psychology, Research Methodology, Research Problems, Russian, *Second Language Learning, Sentence Structure, Teaching Methods, *Traditional Grammar, Translation, Vocabulary Development

Abstract

An experiment in language behavior comparing two methods of learning grammatical word order in a new language presents scientific evidence supporting the use of pattern drills in foreign language teaching. The experiment reviews the performance of three groups attempting to learn small segments of Russian "microlanguage": (1) a drill group learned through pattern drill, (2) a control group was trained with grammar rules and vocabulary lists, and (3) an experimental group learned through a combination of these two methods. Conclusions are discussed from the point of view of the linguist, the language teacher, and the psychologist. A general discussion of grammar and experimental procedures is included. Extensive statistical material, references, and an appendix are also furnished. (RL)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

A Report from -
The Department of Psychology

ED033628

THE LEARNING OF GRAMMAR
AN EXPERIMENTAL STUDY

Jane W. Torrey

Experimental Studies on
the Learning of Language

Progress Report II
Revised

This investigation was supported by
Public Health Service Research Grant No. 07167
National Institute of Mental Health

January, 1966

Reproduction in whole or in part is permitted
for any purpose of the United States Government

FL 001 494

Connecticut College
New London, Conn.

ED033628

Table of Contents

Acknowledgements

Summary

The Learning of Grammar - 1

Psychology and the Learning of Grammar - - - - - - - - - - - - - 9

Hypothesis - 17

Method - 20

Learning Procedure and Apparatus - - - - - - - - - - - - - - - 26

Testing Procedure - 34

Results - 38

Discussion - 47

Conclusions from the viewpoint of the linguist - - - - - - - - 55

Conclusions from the viewpoint of the language teacher - - - - 56

Conclusions from the viewpoint of the psychologist - - - - - - 58

References - 66

Appendix A

Acknowledgments

The most important and most efficient member of the team of workers participating in this experiment was the Research Assistant, Mrs. Jean Ebenholtz. She organized the materials and procedures, served as experimenter for most of the subjects, carried out the statistical analysis and oversaw the preparation of the final report.

In the designing and pretesting of the materials, Miss Lenore Farmer gave most valuable and devoted service. She also served with Miss Elizabeth Dawe, Miss Carolyn Parker, Miss Rowain Schultz and the author as experimenters for some of the subjects. Mrs. Marie Edwards and Mrs. Marcia Siegel helped with the statistical analysis.

My thanks for help and encouragement in the design of the experiment are extended to Professor John B. Carroll and Professor Roger Brown, both of Harvard University and to Dr. Martin D. S. Braine, of Walter Reed Army Institute of Research. I am also grateful for statistical advice from my Colleague at Connecticut College, Dr. Robert L. Rhyne.

I am also very much indebted to Mrs. Julie King for her careful work in preparing the reports.

Summary

The purpose of this experiment has been to compare the effects of two different methods of learning word order in a new language, one based on the pattern drill model and the other on the vocabulary and grammar rule model. The selection of effects to measure was based on the assumption that a "good" method of learning would be one which produced the same types of behavior and perceptual effects as have been shown in the case of the native language. Several conclusions can be drawn at this time.

1) Our version of "pattern drills" enabled our drill subjects, trained exclusively with pattern drills, to translate sentences from written English to spoken Russian faster and more correctly, both in vocabulary and in word order, than did our control subjects, trained by our version of lessons with vocabulary plus grammar rules. Our Experimental Group, whose training included some pattern drills but only vocabulary training on the words used in the actual tests, was better than the controls in number of sentences correct and in word order, but was not significantly faster nor better in vocabulary on the test. The experimental method was significantly inferior in most of these measures to the drill method, which had pattern drills at all stages of training.

2) Drill and experimental subjects showed greater ability to produce Russian sentences, if we use imitation of spoken Russian as a criterion of production. The fact that this "memory span" test differentiated the groups only when the strings were grammatical and not when they were ungrammatical, implies that it is indeed some clearer expectation of word order that accounts for the difference and not merely a superiority in vocabulary. If we consider the English to Russian translation as an instance of production with understanding, then we have already shown that drill facilitates that task

as well. However, when it came to comprehension, measured in our case by the ability to translate spoken Russian into English, we were not able to show much superiority for the drill method. Slight differences favoring the Drill Group over the Control were significant only in the case of the vocabulary measure.

3) No conclusions concerning our hypothesis were drawn from scattered differences in form-class clustering in freely recalled words, though significant clustering did occur in all groups.

4) Both drill and experimental subjects showed a keener grammatical sense than controls in their performance on our modified "Cloze test", which required selecting words of appropriate form-class to fill blanks in Russian sentences. Both the first two groups were faster and drills were also more accurate in this task.

5) Most of the measures used were significantly correlated with the MLAT total scores.

In general, it seems safe to conclude that our experimental test has shown that a pattern drill method of learning grammar produces greater skill than is reached without such drill and that the habits thus acquired resemble in several ways those of the native language. However, it was difficult to account for the results in terms of sequential associative responses and therefore it seems realistic to look for an account of grammar learning in terms of conformity to more abstract rules.

The Learning of Grammar

An Experimental Study

Many people think of learning a new language as being primarily a matter of learning a new set of words to express familiar meanings. While there are always a great many words to learn, vocabulary is only a part of the task. Another of the important problems is to learn to pronounce and to recognize spoken words in another language. This usually means acquiring a new frame of reference for judging and classifying speech sounds. The familiar types of sounds from the native language must be ignored and difficult distinctions attended to. Without a familiarity with the new sound system it is difficult either to speak or understand the spoken word.

The subject of this paper, however, is neither vocabulary nor sound learning, but still a third task that confronts the learner of a new language. Even if a person could pronounce and understand words and even if he also knew their individual meanings, he would still be helpless to use the language without a grasp of sentence structure. In any language there are standard syntactical structures, including expected orders in which words may come. Only certain orders are permitted and the choice of orders may profoundly affect the meaning of what is said even while the words remain the same. This is particularly true of languages such as English and French where there are relatively few grammatical markers attached to words. If we change the English sentence "The man bites the dog" to "The dog bites the man", we change the meaning radically. No one could be said to know English unless he had a good grasp of the implications of these changes in order of words.

The purpose of the present study is to examine experimentally the process of learning grammatical orders, in other words the "features of arrangement" inherent in a language. More specifically, we will be comparing two methods of learning word order in a new language. In making this comparison we are at the same time testing a hypothesis about the nature of the skill being learned since the two methods reflect two different appraisals of the kind of learning required.

According to one view grammar learning consists of understanding and remembering the formal descriptive rules for the formation of sentences. It should follow that a person who knew these rules and knew the words would be able to speak and understand the language with no further practice required. This kind of analysis of language learning is implied in the organization of many text books. Each lesson contains a list of new words and a selection of grammatical rules. The rules are stated technically and illustrated by one or more instances. The rest of the lesson consists essentially of repeated tests to assure that the student can apply the rules to a variety of sentences. These tests are generally labeled "exercises", and it is obvious that the "exercise" is of an intellectual kind. The examples require some thought because both words and rules are relatively unfamiliar. Each successive item will require new words and different rules, frequently several at a time. There is no expectation that the student be able to rattle them off at conversational speed one after another. The planner of the lesson seems not to have wanted the response for any given item to be too easy. If the answer were obvious, the learner would not be exercising his ability to remember and use the rules. Presumably his eventual ability to speak the language would be the result of a high development of this intellectual skill.

Many linguistically oriented teachers of language think that the foregoing implicit analysis of the grammar-learning process is a false one. In the words of Politzer (1960), "linguists distinguish rather unanimously the learning of language (performing in the language) from the learning of rules and grammatical terminology (p. 2)." A practical result of the difference is pointed out by Delattre (1947), "Students who know the rule still fail to apply it as soon as they speak. Every day, we see advanced students who know perfectly well the rule for the use of imperfect after si and who keep on using the conditional instead (p. 242)."

If descriptive rules are different from grammatical skills we may still ask whether they are necessary or useful in the acquisition of a new language. Many language teachers seem to think that the rules play at best a secondary role in the important kind of grammar learning. "From the linguistic point of view, the legitimate objections are not to 'grammar' as such, but rather to the substitution of the learning of grammatical terminology and rules for the learning of the constructions themselves.....(Politzer, 1960, p. 6)." "Rules and grammatical explanation serve the purpose of describing to the student what he is doing and not of prescribing what he ought to do: constructions in the foreign language must be learned as a whole rather than assembled (Politzer, 1960, p. 10)."

Brooks (1964) goes so far as to question the usefulness of any descriptive knowledge of grammar rules. "It is not necessary to take a watch apart in order to tell time, nor to be able to disassemble and reassemble an internal-combustion motor in order to drive a car. Since every speaking person has mastered his own language through imitation and analogy and without benefit of analysis, it stands to reason that something of this ability will aid him in the learning of another language (p. 147)." Another statement by the same author goes even further in rejecting the formal rule as a teaching instrument:

> "If the understanding of a language pattern led immediately and directly to automatic control of that pattern, language learning would be far different from what it is. The intellectual perception and verbal statement of the nature and function of language patterns are of course of great interest from the point of view of scientific description. Whether or not this description is an aid in learning is another matter. In point of fact, the formal verbalization--the rule--may be either a hindrance or a help, and manipulation of the interplay between practice and rule is one of the most delicate operations of the language teacher (Brooks, 1964, p. 153)."

Delattre (1947) clearly opposes the use of rules as means of learning grammar. He would exclude them entirely until after the essential learning is complete.

> "We may say that one who knows the rules still has to form habits of applying them; and since the rules can be a hindrance in forming such habits, their formulation must be postponed until the grammatical habits have been definitely acquired (p. 242)."

> "Wouldn't it be better if they didn't know how to formulate the rule but always practiced correct speech? In the end the only thing that counts is to have correct speech habits. And it would not appear that the best way to acquire such habits is to begin by learning the grammar rules. It even seems that learning the rules retards forming correct speech habits. This is perhaps because the reasoning process becomes a habit in itself and forms an obstacle to the automatic reactions of speech (p. 242)."

If grammatical skills are not learned by an intellectual process, it is logical to suppose that they will be acquired by some other means such as repeated practice of correct forms. Politzer (1960) expresses the point in psychological terms: "Linguists have realized that language is 'behavior' and that behavior can be learned only by inducing the student to 'behave'--in other words to perform in the language (p. 2)." "The actual learning of the foreign language takes place primarily by performance and habit-formation on the part of the student (p. 11)." Fries (1945) is in essential agreement when he says: "The fundamental matters of the language that must be mastered on a production level should, as soon as possible, be made unconscious habits. For this purpose many whole sentences, questions and responses, demand repetition and more repetition and these will become automatic reactions early (p. 8)." "Only after much practice of the same 'patterns' with diverse content do the patterns themselves become productively automatic (p. 9)."

Both Politzer and Brooks go on to compare the need for practice in grammar with that required in motor skills.

> "An analogy may perhaps elucidate the point made above: if we teach someone to drive a car using a manual shift we have to point out to him how to shift gears.....This explanation of the pattern is necessary but no one will ever assume that the explanation alone will teach the student how to shift gears--the only thing that will teach it is the student's performing the act of shifting--correctly and repeatedly. No driving teacher will take it for granted that his student has mastered the skill because he can describe the pattern involved in the act, or even because he has shifted gears a few times correctly. Again, the actual realization of the pattern must be put out of the realm of consciousness if the student is ever to learn to drive a car. We cannot indeed conceive of any driver who every time he shifts gears says to himself in his mind 'I am now putting the left foot on the clutch, I am depressing it, while the other foot is off the accelerator, and my right hand is shifting the gear while the other is steering the car, etc., etc.,' Some driving this would be! Perhaps the driver may pass his test, though it is very doubtful. In the same way the student who says, upon answering a question, 'The noun stays at the head of the sentence, then the pronoun comes, then the verb, etc.,' may perhaps pass his French examination,--especially if it is a written examination that allows lots of time for his responses. But he cannot speak French any more than our driver can drive a car in the street (Politzer, 1960, p. 16)."

The solution, then, is practice. Efficient practice, like driver training, would not be merely driving from place to place but would include selective repetition of the key skills. Brooks, therefore, distinguishes between especially designed drills and attempts at ordinary conversation, which might also be called a kind of "practice".

> "Pattern practice (or structure drill, as it is sometimes called), contrary to dialogue, makes no pretense of being communication. It is to communication what playing scales and arpeggios is to music: exercise in structural dexterity undertaken solely for the sake of practice, in order that performance may become habitual and automatic--as it must be when the mind concentrates on the message rather than on the phenomena that convey it (Brooks, 1964, p. 146)."

If it is agreed that learning of rules is inadequate and that repetitive drills are necessary for a student to acquire grammatical habits, the question remains what sort of drills will accomplish the purpose. The authors already cited agree that drills must represent complete and correct constructions in the language, and they should be selected and arranged in accordance with linguistically valid descriptive characteristics. In particular, says Brooks (1964, p. 153)

> "Instead of asking the student to learn by working with utterances in which a considerable number of variables, none of which he knows very well, are all varying at once, we ask him rather to work with utterances in which there is either identity or minimal change, often even in the same place in the sentence, so that he may become habituated to what is constant and what varies."

To say anything more definite about the methods of constructing learning materials, it is important to specify more exactly the content of this proposed learning. Politzer (1960) compares it to learning how to fit elements into a pattern.

> "The student must not only learn a construction--he must also realize how this construction is 'made up', how it 'comes apart', how some building stones can be replaced by others (p. 6)."

> "In a 'linguistic' teaching approach the construction in the foreign language is the starting point of instruction. The student learns how the construction is made up by exercises in which building stones are replaced by others. This shows him how the construction fits together and what the value of each building stone is. In a sentence like:

> Je veux que vous appreniez le français
>
> we show the student how _français_ can be replaced by _grec_, _latin_, etc., _appreniez_ by _étudiez_, _sachiez_, _compreniez_, etc., _veux_ by _exige_, _doute_, etc. This teaches the student not only the fact that the building stones _appreniez_, _sachiez_, etc., or _veux_, _exige_, belong to the same category since they can fit into the same spot of the construction, but it teaches also the construction, the 'pattern' itself. For while we are replacing the individual elements of the construction by others, the construction itself, the 'pattern' remains constant (p. 7)."

Brooks perceives the effects of pattern drill not so much as the learning of a particular content as laying the groundwork for transfer of training from the drill sentences to others the student will want to produce. He distinguishes "analysis", or stating of linguistic rules, from "analogy", the reapplication of a pattern to new material.

> "A principal reason for the recent popularity of pattern practice is that analogy and analysis as factors in the acquisition of another language have been reassessed. Instead of relying exclusively upon analysis, as we have been doing for centuries in the study of all foreign languages, we now invoke the aid of analogy, which may be defined for our purposes as hidden sameness. Since as children we learn the mother tongue quite by analogy and not at all by analysis, why should we not try to make analogy work for us in the learning of a second language? (Brooks, 1964, p. 152-3)."

> "Pattern practice capitalizes on the mind's capacity to perceive identity of structure when there is a difference in content and its quickness to learn by analogy. Analysis is important in its proper sphere, but analogy is used instead through pattern practice to produce a control of language structure without the time and effort required for grammatical explanations (p. 146-7)."

The object of a pattern drill is to teach grammatical forms for sentences. It does this by somehow inducing the student to produce sentence after sentence in the same grammatical pattern with the successive sentences differing in content but not in grammatical form. There are two basic types of such drills, the transformation drill and the substitution drill. Both are devices for telling the subject what words to use in the sentence pattern he is repeating. In a transformation drill the stimulus is a sentence of a different pattern, such as a question to be answered or a declaration to be converted into a question. In a substitution drill the simulus consists of words to be used in the next sentence. A typical substitution drill in English would run like this:

Teacher: They are at the window.
Students: They are at the window.
Teacher: Theater
Students: They are at the theater.
Teacher: We
Students: We are at the theater.
Teacher: In
Students: We are in the theater.
etc.

Two important points should be understood about such a drill. The first is that although students seem to be responding to and learning words, vocabulary is not the primary purpose of the drill. The important thing is that the student is getting used to putting these words in a certain grammatical pattern. It is the pattern that is repeated from response to response, not the words, and the pattern is therefore the thing being drilled and learned by such an exercise. The second point is that the performance of the pattern becomes automatic and does not demand the student's attention, as it should not in his actual use of the language. The changes in words are supposed to occupy his conscious thoughts while the grammar is unconsciously determined by habit. The theory is that grammar is not "learned" until the student is able to produce sentences that accord with the rules while thinking about something else.

It is the purpose of the present experiment to develop laboratory paradigms for language lessons based on pattern drill and for lessons based on vocabulary plus grammar rules, and then to compare the performance of subjects using the two methods in some of the tasks in which they might be expected to differ. Before describing our experimental techniques in detail, however, it is worthwhile to give more attention to some of the psychological experiments which have already been done on acquisition and effects of linguistic habits or language-like habits.

Psychology and the Learning of Grammar

If we accept the notion that grammatical sequences can be learned as automatic unconscious habits, the question for the learning psychologist remains "what is the nature of the habits that make grammatical speech possible?" In search of an answer we can consider studies both of the acquisition of grammar or grammar-like habits and of the effects of grammatical habits on perception, memory and behavior.

Acquisition

Contextual generalization. As a first step it should be pointed out that to learn word order is not the same as to learn a sequence of particular words. A sentence frame is a much more general pattern into which a tremendous number of different words can be put. Braine (1963) has suggested that to learn word order may be to acquire associations between words and their possible positions in phrases and sentences rather than between words and other words. He calls this learning "contextual generalization" because the basis of generalization, or substitutability of one word for another in a sentence, is not an intrinsic property of the words, but rather of the contexts in which they occur.

Braine did an experiment to show that associations are possible between words and positions in sentences apart from the particular other words that make up the sentences. The material for one of his experiments was an artificial language consisting of two kinds of words. A-words and P-words, and one kind of sentence, consisting of an A-word followed by a P-word. The total vocabulary of the language was as follows:

A-words	P-words
kiv	bew
juf	mub
foj	yag

Braine's subjects, third graders, were given substitution problems like this:

kiv____________________(juf, bew)

The subject's task was to select from the two items in parenthesis the one which could go in the blank. (The correct response is "bew", a P-word following the A-word "kiv".) The subject was informed whether he was correct and was presented several of the possible problems involving the first two words in each set repeatedly in random order until seven in succession were answered correctly. This completed the learning phase. The test was for generalization of the context, that is, it tested whether the correct position of a word had been learned regardless of what particular other words were used in the sentence. The test problems introduced words not used in training, for example:

foj____________________(kiv, mub)

Subjects had had training in placing "kiv" in the first position and "mub" in second position, but had not yet seen the word "foj". Since they did better than chance in these generalization problems, in this case in selecting "mub", it was concluded that an association had been formed, not between two words, but between a sentence position and a word.

An interesting and perhaps significant finding of Braine was that his subjects, while selecting the correct word, could not usually tell in words why it was correct, that is, they could not verbalize the rule that said "mub" goes in second position. They commonly did not realize that the generalization sentences were new. Their explanations of their choices included such answers as "It sounded right." This sounds very much like the way we identify correct grammar in our own language. It is as if they had acquired the "unconscious habits" so sought after by language teachers with their drills.

"Place" Learning. A further support for the possibility of learning the "locations" of words in a sequence apart from the formation of associations between particular words comes from an experiment of Ebenholtz (1963). He compared the learning of serial lists presented in two different ways. The first was the usual way, with the same list in the same order being presented in each practice trial. In the second method, the same items were learned in the same serial order but for each trial the list was begun at a different point. Thus a given syllable always followed and preceded the same other syllables, but on successive trials it might occur second, third, or in any other list position. If serial learning is just a learning of sequential associations, then learning under the two conditions should have required an equal number of trials. Instead, the method that kept each syllable in the same absolute position throughout produced easier learning. Ebenholtz concluded that location in relation to the beginning and end of the list was part of what was being learned.

When it did not remain constant, serial learning was delayed. In another experiment Ebenholtz also found that transfer of learning from a first to a second list was greater for items that occurred in both lists in the same positions relative to the beginning and end of the list. Items repeated in the second list but coming in different list locations showed less transfer.

Although Ebenholtz was using a serial learning technique, not a choice procedure like Braine's, it may well be that the nature of the learning produced was somewhat similar.

Effects of Grammatical Habits

It may also be interesting to consider some of the effects of linguistic habits shown in experiments that deal with habits built up through normal use of the native language. In this way we can examine some of the less obvious outcomes that may be expected if, in a second language, the same type of linguistic habits are acquired as in the first.

An important difference between the experiments of Braine and Ebenholtz and the studies that follow is that the former used nonsense materials. All natural languages have meaning and it seems intuitively as though meaning must contribute to grammar learning. Thus in turning to studies of meaningful language we may introduce a much more complex learning process. The following studies have not attempted to bring about learning experimentally, but have simply taken for granted the normal processes of the native language learning and measured various aspects of the habit systems that have resulted.

Sequence of Skills in First Language Learning. Fraser, Bellugi and Brown (1963) have devised some interesting techniques for assessing various kinds of developing first language skills in children. In an attempt to answer the question whether a child learns to understand grammatical signals before he can produce them himself in his own speech, they found it necessary to distinguish between rote imitation and production with understanding of what is said. They found children who were able to imitate a grammatical sentence but unable to comprehend the same sentence. Comprehension was tested by presenting two pictures one of which was accurately described by the test sentences and the other similar but not corresponding to the sentence. If the subject could point to the right picture he was said to have understood the sentence. The distinctions represented by the difference between the pictures were all grammatical ones. For example, a pair of pictures was prepared, one showing two sheep jumping over a fence and the other showing one sheep jumping and another just standing. The child was shown the two pictures and the experimenter said that one picture was called "The sheep is jumping," and the other was called "The sheep are jumping." The experimenter did not say which picture was which but asked the subject to point to "The sheep are jumping." If he pointed to the one with two sheep jumping, he was credited with comprehension of the distinction between "is" and "are". In order to test imitation the experimenter simply pronounced one of the sentences and asked the subject to pronounce it after him. There was no way to know from this whether the subject understood what he was saying. However, the experimenters were able to test production with comprehension using the same pictures. This time after the pictures were shown and the names pronounced, the experimenter himself pointed to one of the pictures and asked the subject to name it. Now if the subject correctly said "is" or "are", it was concluded that he also understood the significance of the difference.

The results of the study showed that imitation was more likely to be correctly performed than comprehension and comprehension than understanding production. This result is understandable in view of the fact that understanding production in a sense requires both the abilities separately tested in the two other procedures. It also established, if there was any question, that meaning requires a separate and different learning process from that needed to produce correct sequences of sound. Although Fraser, Bellugi and Brown were not studying the learning process as such, their concepts and measures are of some interest because they give us some ideas about what to expect as language habits begin to develop.

Intelligibility. Miller & Isard (1963) reported that sequences of words in grammatical order were more readily apprehended through a barrier of noise than the same set of words presented in non-grammatical order. The subjects' task was to listen to the words spoken in fairly rapid succession through earphones and to repeat them in order aloud into a recorder while simultaneously listening to the words that came next. Grammatical sequences enabled them to do the job more efficiently. This result is interpreted to mean that they have somehow acquired habits which organize the perception of words into more efficient units.

Memory. Although in the preceding experiment Miller and Isard were thinking of the task primarily as one of perception over varying amounts of noise, the subject also had to retain the words in mind long enough to say them while listening for the next ones. Some memory is involved and to that extent the task partly resembles that of a test of memory span, of repeating words in order after just hearing them.

Psychologists have shown in a variety of ways that memory as well as perception is affected by the language habits of the subjects. Miller and Selfridge (1950) had subjects listen to sequences of words which varied according to their statistical approximation to English word order. The percent of words which the subject was able to write down after hearing the sequence varied according to the degree of approximation to English. Epstein (1961) compared the number of trials required to repeat correctly strings of nonsense syllables to which there either were or were not added inflectional suffixes to produce a kind of Jabberwocky sentence. The same comparison was made with words which either were or were not placed in grammatical order. The difference in trials to criterion significantly favored the strings with "grammatical" structure. Marks and Miller (1964) had subjects listen to sequences of 25 English words arranged either in five sentences or in five, 5-word, non-grammatical strings. Subjects were able to write more strings completely correctly if they had been arranged in sentences. All of these findings imply that grammatical sequences enable the subject to form larger mental units than single words and thus retain more material.

Sensitivity to form-class. Another type of behavior that presumably somehow reflects linguistic habits is word association. Ervin (1961) has pointed out the shift from syntactic to paradigmatic responses in word association between kindergarten and the sixth grade. She attributes the paradigmatic response to an indirect association mediated by the common preceding verbal context in which speakers of English will have experienced words of the same form-class. Brown and Berko (1960) have shown that the change in type of associations parallels the acquisition of ability to identify form-class of "words" from the sentence context in which they are presented, thus confirming the connection between the pattern of word associations and developing grammatical habits.

If the consequence of grammatical habit is the ability to discriminate words which have different privileges of occurrence in sentences, then the more firmly these habits are developed, the more decisively should words be differentiated. This implicit classification of words might be expected to influence the pattern of sequential production of words, such as occurs in free recall. Bousfield (1953) has demonstrated a tendency for words produced in free recall to be said in a nonrandom order with respect to different classes of words. Specifically, when the subjects who have been once exposed to a list of words which is readily divided into classes are then asked to recall them, they seem to recall members of the same class sequentially more often than chance would predict. Words of a kind seem to "cluster" together. Although Bousfield's experiments tend to be with lists that are classified with respect to referential meaning, it is not impossible that words in a language lesson might become classified, as habits of use develop, with respect to form-class and thus be clustered on that basis in free recall. Those with stronger grammatical habits would be expected to show greater clustering with respect to form-class.

Readability. Taylor (1953) proposed a measure of readability of written material based on the idea that readers are partially able to anticipate what words are coming up from what has gone before. If in a given text words are easily anticipated, reading would be less difficult. The assessment of texts is not of particular interest to us here, but the technique and its use indicate the presence of measurable habits which are determined by knowledge of language. The measure consists of deleting words at regular intervals and asking subjects to guess what words have been left out. The percentage of correct guesses for the text measures its readability. Anything other than purely random guessing, on the part of subjects, would indicate an influence of language habits. We could turn the question around and assess the strength of the habits in subjects whose skills with the language were not yet highly developed. Presumably as they become more skilled in a language they would be better able to fill the blanks accurately with respect to grammatical form-class.

Hypothesis. The hypothesis that was tested by the present experiment can be stated either from the viewpoint of the language teacher, as an evaluation of the pattern drill, or from the viewpoint of the psychologist, as an assertion that grammatical word order can be learned independently of any sequential associations between words. Subjects trained by the pattern drill method were compared with subjects trained by a grammar and vocabulary method. The comparison tests measured both speed and accuracy in producing correct grammatical sentences in the new language. In addition, measures were made of grammatical habits like those that have been demonstrated for the native language. It was predicted that pattern drills would result in greater fluency, or speed, and more accurate word order, but not necessarily greater accuracy in vocabulary. It was also expected that the other measures of habit in language would show a greater development in subjects who learned by pattern drills. Memory span would be increased for grammatical sequences of words in the new language. In a modified "Cloze Test", more grammatically appropriate words would be chosen to fill blanks in incomplete sentences. In free recall of words from a new vocabulary there would be greater clustering according to form-class. Perception of foreign-word sequences, measured by "imitation" in the memory span test, would also be measured by a test requiring comprehension, a test in which the response had to be an English translation. If grammatical habits contribute to the ability to form relevant "chunks" from strings of words, the ability to comprehend and to translate should also be increased. All such changes in language habits were expected to be greater for subjects whose training included pattern drills than for subjects given only the appropriate grammatical rules and vocabulary.

In order to test the psychological hypothesis that grammatical habits are not associational sequences of particular words, it was necessary that training and test materials be so designed that at no time in the training did subjects learn any sequential associations between the words to be used in the tests. Since test words had to be learned somehow, they were introduced in the training one at a time. In the pattern drills, the test words were presented in sentences, but no one training sentence ever needed to contain more than one test word. The other words in the sentences were for training only and never occurred in the tests.

The experimental design used three groups of subjects. One group, called the Drill Group, had all its training with complete sentences selected to represent three different grammatical patterns. The initial examples were made up of training words (later called "x" words) with the test words (called "y" words) introduced one at a time in the latter part of the training. Thus all the training was with complete sentence patterns, but test words were never presented in sequence with one another.

Another group, called the Control Group, had training based upon the grammar and vocabulary method. Grammar rules were presented with illustrations and exercises in their application. Words were learned one at a time as in a vocabulary list.

Still another group, called the Experimental Group, combined the two methods in such a way that pattern drills were formed using only the training words and the test words were learned entirely by the vocabulary method. The purpose of this group was to give a somewhat tighter test to the hypothesis that pattern drill provides some kind of skill with the abstract pattern sequence of word-types rather than any particular sequences of words. If the pattern really were somehow apprehended apart from any particular instance with particular words, then once it was learned, it should have been possible to substitute new words in appropriate slots regardless of the particular way in which the new words were learned. The drill group had learned test words in sentence contexts and even though all test words were carefully separated from one another in the training sentences, it was nevertheless possible that differences in their ability to form sentences combining these words might be due primarily to the fact that the words themselves had been directly associated with sentence locations. If this were so, the pattern drills on the other words might be having little or no effect, that is, no abstract learning of a pattern of word-types might be involved at all. If, on the other hand, pattern drills produced some more abstract kind of learning of sentence patterns apart from the words that made them up, then it should be possible to acquire the sequence patterns quite separately from the words to be used in them. In this case our Experimental Group should perform as well as the Drill Group.

Method

An experimental study of the development of grammatical skills and habits under different conditions of learning requires, first, a set of materials that resembles as much as possible a second language; second, experimental procedures which duplicate the essential features of the learning conditions to be compared; and third, a set of tests that measure the skills and habits to be expected in language learning.

Materials. Our subjects learned a small segment of an actual foreign language, Russian. Subjects were selected who had no prior experience with Russian. Our small segment, a kind of "microlanguage", contained three types of sentences and a total of 11 form-classes of words. For each form-class, there were 6 words, a total lexicon of 66 words. This "microlanguage" differed in several ways from the whole Russian language. Obviously it was much smaller and simpler. There were no cases or tenses, since no words occurred in more than one form. The form-classes of the "microlanguage" were all different, that is, a given word could occur in only one place in one kind of sentence. (In the Russian language as a whole several of our form-classes would be considered the same, that is, some words could have been substituted in other places in other sentences and still produced correct Russian.) None of our sentences was a transformation of any other, so that all had to be learned entirely separately. In addition, word order was fixed in our language, whereas in real Russian other orders of these sentences are possible. Within its limitations, however, our "microlanguage" closely resembled Russian. It used Russian sounds, Russian words and correct Russian sentences.

Not only was our material a selection of many possible elements of a whole language, but the kinds of practice and performance used in the experiment were limited compared with the whole variety of uses to which a language can be put. In learning and in most of the tests our subjects were reading or translating from typewritten stimuli in Russian or English and making spoken Russian responses.

Since it would have been possible to do the present experiment with still more limited artificial materials, we should offer some reason for burdening ourselves and our subjects even with so many difficulties of a real language, some of which were irrelevant to the specific issue of learning word order. The use of unfamiliar sounds and words that have meaning introduced complications. We felt that it was justified primarily in the name of realism, since any student of a new language will have to struggle with these elements at the same time he is learning new grammatical systems. Particularly in the case of meaning, it seems as though the added knowledge might have an important effect on the manner of learning grammar.

It might also have been easier to deal entirely in reading and writing, as Braine did. We chose to use exclusively oral responses in order to be surer we were dealing with the same processes people use to learn languages. Language is primarily a spoken phenomenon. Writing is both secondary to and different from the spoken language. Generalizing from written to oral grammar learning would be questionable at best.

We chose to work primarily with written stimuli and with translation responses because their convenience and the control they offered seemed to outweigh the limitations. Pictures would have been much better stimuli in many ways, but the meanings that can be easily conveyed by them are very few. Written stimuli were easier for subjects to use than spoken ones because they could be kept before their eyes as long as necessary. We also judged that typewritten English words would convey meaning as well as the spoken word. Translation rather than some other response was chosen because it gave a clear-cut task for learning and a means of identifying correct responses.

The reason for using typewritten Russian words as stimuli requires a different kind of explanation. For this it was necessary to take the time to teach the subject to read 66 words in Russian. Although this learning was not directly used in any test except the "Cloze", we retained it because of its value in teaching the sound system of Russian. If strange sounds are to be combined into words, at least some new notations are very helpful. The Russian alphabet seemed, on balance, to be the most convenient system. Preliminary studies using purely oral training in Russian words had indicated that without written stimuli subjects had much more difficulty with learning to say and understand the words and that in the end as much time was lost as would have been required for reading training. Our method also separated the task of learning sounds from those of meaning and grammar. Problems of reading and pronouncing were well in hand by the time the elements important to the experiment were introduced. (Another advantage of alphabet training was that our subjects valued the knowledge. The ability to "read" Russian was almost as much incentive to them as the one dollar or so per hour they were paid.)

A final complication in our material has been the use of more than one kind of sentence and a considerable variety of words. Braine's study used only one kind of sentence and therefore fewer words. A preliminary study had shown, however, that a single 3-word sentence pattern with six possible words for each position was too easy. There was no difference in performance between drill and control subjects and both responded almost as fast as speech would allow. Probably one of the limitations of grammar rules as crutches of speaking is that it is necessary to locate the right rule from a large repertoire. The same is probably true of vocabulary learning, that is, it is excessively easy to pick out the right word when only 18 words and 3 form-classes are known. We felt, therefore, that three kinds of sentences and 66 words would go a long way toward making the complexity of the task equal to that of a few lessons in a real language.

Figure 1 shows part of the material, one of the sentence patterns used with the 18 words which the subject learned. Each sentence contains three "words" or slots. In this set the first slot can be filled by an infinitive, the second by a pronoun and the third by an adverb. Six words are provided for each slot, any one of which may be used with a word for each other slot so as to produce a reasonable not-too-anomalous sentence. Approximate English translations of the words separately and of the sentence as a whole are given. The sentence type is an example of a grammatical pattern that might be taught by means of a pattern drill. The order of words and structure of the sentence is quite different from that of normal English, so an English-speaking person would be learning a pattern quite unfamiliar to him. This "corpuscle" of language material is useful for our purpose, because it provides for a large variety of different sentences, all of the same grammatical pattern with a limited variety of words to be learned.

Six such sets were constructed. Each subject learned three of these. One of the three was a 3-word pattern in which the Russian grammar differed markedly from that of the English translation. One was a 4-word pattern in which the Russian closely resembled the English word order. The third was a 4-word pattern with different word order from English. Since the subject learned 6 words for each slot, she learned 18 words of three different form-classes for the first sentence frame and 24 words of four different form-classes each for the other two sentence frames, a total of 66 words.

Two complete "corpuscles" of material were used, each answering the above description, but using a different set of 3 sentence patterns and a different set of 66 words. One set was designated "A" and the other "B". There is nothing in the hypothesis to predict a difference between these two sets. They were used to obtain a partial control over the effects of particular patterns or words.

Subjects. Undergraduate students at Connecticut College served as subjects. All were women and none had any knowledge of Russian. They were paid approximately one dollar per hour for the 15 hours they served. The Modern Language Aptitude Test (Carroll & Sapon, 1959) was given in the first session and subjects were assigned to experimental groups so as to equalize language aptitude scores as much as possible Table 1 shows the individual and mean scores on the MLAT for all subjects.

x	КУРИ́ТЬ	ВАМ	НЕЛЬЗЯ́
	to smoke	for you	not allowed
	You are not allowed to smoke.		
x	СЧИТА́ТЬ	МНЕ	МО́ЖНО
	to count	for me	possible
	It is possible for me to count.		
x	ПРА́ВИТЬ	ЕЙ	ТРУ́ДНО
	to drive	for her	difficult
	It is difficult for her to drive.		
y	ХОДИ́ТЬ	ИМ	НЕВОЗМО́ЖНО
	to walk	for them	impossible
	It is impossible for them to walk.		
y	УБЕЖА́ТЬ	ЕМУ́	ЛЕГКО́
	to run away	for him	easy
	It is easy for him to run away.		
y	СПАТЬ	НАМ	НУ́ЖНО
	to sleep	for us	necessary
	It is necessary for us to sleep.		

Sample 3-word Sentence Pattern

Fig. 1

TABLE 1

Individual MLAT Scores

Group	Drill		Experimental		Control	
Material	A	B	A	B	A	B
	123	142	120	146	126	164
	153	152	117	110	145	121
	141	154	120	131	141	156
	159	122	141	122	154	157
	109	129	161	135	155	103
	149	148	162	156	142	164
	129	139	142	150	123	121
	148	119	152	163	122	144
Group Means	139	138	139	139	138	141

Figures 2 and 3 show the distributions of raw scores and percentiles on the total MLAT as well as raw scores on the five separate parts of the test. It is clear that our subjects' scores were not normally distributed on the test as a whole nor on Parts I, II and V. On the total scores, we had many very high scorers, perhaps explained by the fact that the subjects volunteered in the knowledge that the experiment dealt with language learning. A number of them were language majors, though none, of course, Russian majors since anyone who knew Russian at all was excluded. A curious fact, however, is that so few scores are between the percentiles 65 and 85 where most of the scores would be expected to fall.

Experimenters. Each subject was trained and tested by the same experimenter throughout. Experimenters were students of Russian with sufficient command of the language to handle the material of the experiment. None was a native speaker of Russian but all were able to make the necessary distinctions between most Russian sounds and their nearest English equivalents. Subjects trained by them compared favorably with average beginning students in their pronunciation. Live experimenters were chosen rather than recordings of native voices because of the greater ease of adjusting them to the individual requirements of the subjects.

Learning Procedure and Apparatus

Preliminary training. Before beginning the actual language learning, each subject had preliminary training in the Russian alphabet and sound system. Letters and syllables in Russian characters were typewritten on cash register tape and mounted in an Answer-Mate Attachment for a TMI-Grolier Min-Max II Teaching Machine. By turning a knob the subject could present to herself one letter or syllable at a time. Each new letter was pronounced first by the experimenter, then by the subject and then again by the experimenter. The subject was asked to repeat if her pronunciation was notably different from that of the experimenter. After a few presentations, the subject was asked to respond first and was then either corrected or confirmed by the experimenter. The letters were presented a few at a time with the vowels first, appearing singly. After the vowels were fairly well learned syllables containing the various consonants and the easier vowels were learned a few at a time, starting with the easier sounds and working up to combinations of the more difficult vowels and consonants. Next, the subject was drilled on a series of more difficult syllables taken from the words in the material to be used. Finally, the actual words of the experiment were presented and drilled until the subject could read each one aloud

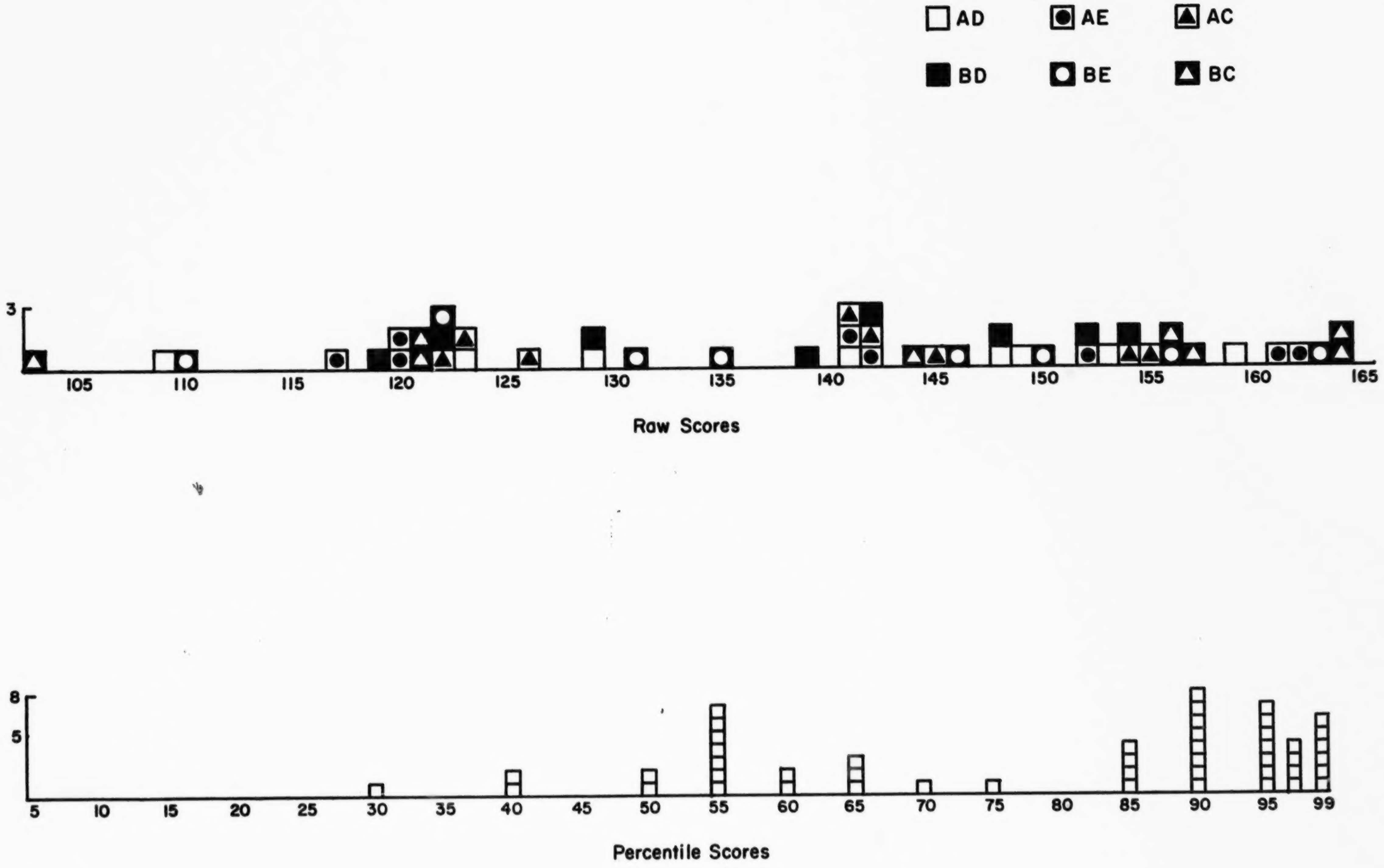

Fig. 2 MLAT Scores

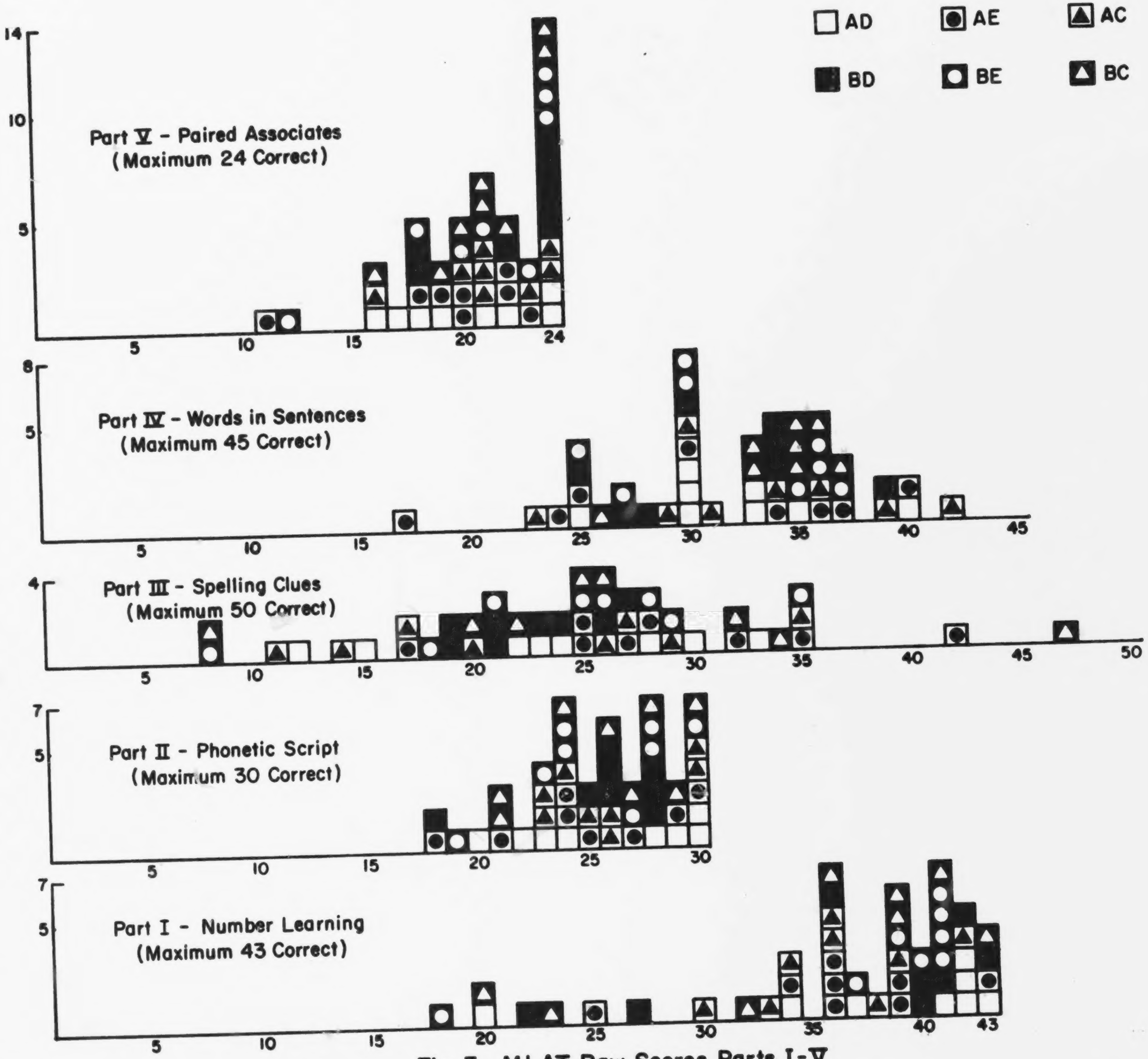

Fig. 3 MLAT Raw Scores Parts I-V

at a glance. Up to this point the subjects had no idea of the meaning of the words. None of the words was an obvious cognate to any English word.

Grammar Lesson. After the preliminary reading and pronunciation practice, the subjects were ready to start learning the experimental microlanguage. Before training, all subjects were given a short "grammar" lesson which is shown in Fig. 4. The sentences were read and translated into English sentences. There was a brief descriptive statement of the sentence structure, including a word by word translation into English. The lesson covered the first three sentences in each set to be learned containing three words for each slot hereafter called the "x" words. No mention was made in the grammar lesson of the last three sentences nor of any words used in them. These other words, called the "y" words, were for use in the testing. The "x" sentences in Russian and English and the word-by-word translations mentioned in the lesson were read and simultaneously presented in writing on the Min-Max II Teaching machine so that the subject saw all the "x" material just as it appears in Fig. 1 while it was being read and discussed.

Procedure for the Drill Group. The Drill Group was to learn the microlanguage by a method as close as possible to pattern drills used by language teachers. After the grammar lesson the subject was given the same teaching machine and instructed to turn the knob herself. The initial frames are shown in Fig. 5. First came a frame in which a complete Russian sentence appeared with an English translation of each individual word typed under it. The subject was instructed to respond to this and all other frames by saying a complete Russian sentence. At first the Russian words could simply be read, but she was also instructed to note the English because she would be expected later to say the Russian when only the English was given. In subsequent frames the Russian was omitted, first for one word at a time and finally for all words so that the subject gave the entire Russian sentence from memory with English words as stimuli. Then other words from the "x" set were introduced one at a time, first with both Russian and English given and then with English only. However, the correct response to each frame was an entire Russian sentence of the pattern to be learned with no more than one new word in each frame. After the first 27 frames all the Russian words had been presented and were no longer seen, so that subsequent frames had only strings of English words to be responded to with Russian sentences. Although most subjects had learned the Russian words fairly well by this point, they were still provided with further training because the experimenter throughout the learning phase of the experiment always pronounced each complete sentence after the subject. If the subject had made an error, she was to repeat the sentence correctly and listen for the experimenter to repeat it correctly again. The subject was instructed not to turn the knob for the next sentence until both she and the experimenter had said each sentence correctly. Pronunciation at this time was corrected only if it could have led to confusion about which word was being said. The subject proceeded through the lesson at her own speed but was encouraged to go as fast as possible.

After a total of 45 frames in which each of the three "x" words had been included 15 times, the "y" words were introduced one at a time, first with Russian and then without. No "y" word ever appeared in the same sentence with any other "y" word, so that the subject never had any practice with any sequences of these particular words. In the next 117 sentences

Grammar Lesson for Pattern B II

"The Russian sentence Кури́ть вам нельзя́. is translated by the English sentence 'You are not allowed to smoke'."

кури́ть means 'to smoke'.

вам means 'for you'. It is a form of the pronoun.

нельзя́ means 'not allowed'.

"The Russian words come in a different order from the way the corresponding words would come in an English sentence. In Russian, furthermore, there is no separate verb to correspond with 'are' in English."

"In English you would not say 'To smoke for you not allowed.' but in Russian you may say Кури́ть вам нельзя́."

"The Russian sentence Счита́ть мне мо́жно. is translated by the English sentence 'It is possible for me to count'."

счита́ть means 'to count'.

мне means 'me'.

мо́жно means 'possible'.

"In this Russian sentence there are no words to correspond with the English 'it is'."

"In English you would not say 'To count for me possible.' but in Russian you may say Счита́ть мне мо́жно. "

"The Russian sentence Пра́вить ей тру́дно. is translated by the English 'It is difficult for her to drive'."

пра́вить means 'to drive'.

ей means 'for her'.

тру́дно means 'difficult'.

"In English you would not say 'To drive for her difficult.' but in Russian you may say Пра́вить ей тру́дно. "

Sample Grammar Lesson

Fig. 4

101	КУРИ́ТЬ to smoke	ВАМ for you	НЕЛЬЗЯ́ not allowed
102	to smoke	ВАМ for you	НЕЛЬЗЯ́ not allowed
103	КУРИ́ТЬ to smoke	for you	НЕЛЬЗЯ́ not allowed
104	to smoke	for you	НЕЛЬЗЯ́ not allowed
105	to smoke	for you	not allowed
106	СЧИТА́ТЬ to count	for you	not allowed
107	СЧИТА́ТЬ to count	МНЕ for me	not allowed
108	to count	МНЕ for me	МО́ЖНО possible
109	to count	for me	МО́ЖНО possible
110	to count	for me	possible
111	ПРА́ВИТЬ to drive	for me	possible
112	ПРА́ВИТЬ to drive	ЕЙ for her	possible

Sample Showing Initial Frames

for Drill Group

Fig. 5

there were always two "x" words and one "y" word so that each "x" word was exposed 26 more times and each "y" word 13 more times.

The next step was to teach the subject to respond with a Russian sentence to a complete and correct English sentence. The stimuli so far were all strings of English words that made a kind of sense but were not in a normal English order. Figure 6 shows the sequence of frames used to introduce the grammatical English as stimulus material. At first the subject was asked to read aloud a correct English sentence after responding to the word-for-word English with correct Russian. Gradually, with the help of both Russian and English written prompting, she was shown how to respond with correct Russian to correct English even though the English words did not now come in the same order as the corresponding Russian words. Drill proceeded for a total of 183 frames with such sentences, giving each "x" word 44 more exposures and each "y" word 17 more exposures.

After the 3-word sentence had been drilled for two experimental sessions, the 4-word sentence which resembled English word order was learned in the same way for three sessions. More examples were required in order to balance the number of times each word was presented, but otherwise the procedure was the same. After a total of 60 frames in which each of the four "x" words had been included 20 times, the "y" words were introduced one at a time, first with Russian and thereafter without. No "y" word ever appeared in the same sentence with any other "y" word. In the next 180 sentences there were always three "x" words and one "y" word so that each "x" word was exposed 45 more times and each "y" word 15 more times. In the grammatical practice that followed each "x" word was exposed 24 times and each "y" word 8 times. Interspersed with this additional practice on the 4-word sentences were more 3-word sentences from the previously learned pattern, giving 16 more presentations of the "x" words for the 3-word pattern and 8 more presentations of the "y" words. In this way the entire contents of all the lessons to date was reviewed.

Finally, the second 4-word sentence pattern was taught in three sessions by the same procedure. This pattern involved a word order different from that of the English translations. The "x" and "y" words for this set were given the same number of exposures in the same procedure as outlined above for the other 4-word set. However, this time in the final session sentences and words from all three sets were interspersed with one another so that the subject saw each "x" word for each of the two 4-word patterns another 24 times and each "y" word 8 more times. For the 3-word set the "x" words were seen 16 times and the "y" words 8 times.

The total training on the sentences required eight 50-minute sessions. Two sessions were devoted to the 3-word sentence pattern and three sessions to each of the 4-word sentences including reviews of previous patterns. It has already been indicated that the grammar lesson preceded training on each set. At the beginning of each new sentence pattern, that is, in sessions 1, 3 and 6, the Russian words were shown to all subjects for review in reading and pronunciation. Each word was presented in Russian twice on the review tape. At the beginning of the second and third sessions for each pattern, that is, in sessions 2, 4, 5, 7 and 8, subjects were given a review of all the individual word meanings to be used in that session. For review sessions 5 and 8 this included two or three sets of words. The English translation of each word was presented twice on each tape and the subject was to respond with the Russian. In addition, at the beginning of the final session

501 to smoke for you not allowed

502 (First read:) You are not allowed to smoke.
КУРИ́ТЬ ВАМ НЕЛЬЗЯ́

503 (First read:) You are not allowed to smoke.

to smoke for you not allowed

504 (First read:) You are not allowed to smoke.
КУРИ́ТЬ

505 (First read:) You are not allowed to smoke.

506 You are not allowed to smoke.

507 to count for me possible

508 (First read:) It is possible for me to count.
СЧИТА́ТЬ МНЕ МО́ЖНО

509 (First read:) It is possible for me to count.

to count for me possible

Sample Showing Introduction of Grammatical English

Fig. 6

for each set, the grammar lessons for all material in the set were presented again. At the second presentation of each lesson the subject was required to translate correctly at least one unfamiliar sentence into Russian to show that she understood the grammar lesson well enough to apply it. For the drill subjects this task should have been very easy since all of their practice was with the same kind of task, but for the control and experimental subjects it may have been more of an intellectual challenge. However, all subjects succeeded in producing at least one correct sentence other than those demonstrated.

Procedure for Control Group. The Control Group learned the microlanguage by a method as close as possible to the common pattern of language lessons in which there is a vocabulary list and a few grammar rules. As with the drill subjects, the grammar lesson was presented first, introducing also all the "x" words. The subject was then told that she would learn a set of words to be used in sentences like these. The words were presented in Russian with the same English words used at first for drill subjects typed underneath. After the first presentation only the English was given and the subject responded with a single Russian word. All the Russian words were presented in the appropriate form for that sentence. The distinctions of case and number therefore were not known to the subject since she saw each word only in the form in which she would use it. The words were presented on cash register tape in an Answer-Mate Attachment. The subject was allowed to proceed at her own speed but was encouraged to go as fast as she could.

Each word was presented to the control subjects exactly as many times as it was to the drill subjects. The sessions proceeded in the same way, with "x" words learned first and then "y" words. The same Grammar Lessons and tests were introduced at corresponding points in the sequence of 8 sessions.

Procedure for the Experimental Group. A third group learned by a method which combined drill training (for the "x" words) with control, or vocabulary, training (for the "y" words.) The procedure for the Experimental Group was exactly like that for the Drill Group up to the point where the "y" words were first introduced. Then instructions were given for the vocabulary method, just like that used for the Control Group. The numbers of presentations of each word in each daily session were matched exactly with those for the other two groups. The principal difference for the Experimental Group was that the "x" and "y" words could never be intermingled in any one training sequence. However, some attempt was made to keep the sequence of stages of learning similar to that of the other groups by fairly frequent switching back and forth from the drill procedures on the teaching machine to the control procedure on the Answer-Mate Attachment.

Comparison of Learning Times for Drill, Experimental and Control Groups. Although the number of presentations of each word was carefully equated for the three groups, the time required to complete the lessons was not exactly equal. Table 2 shows total time actually spent in training by the six groups (three for each set of materials) and also the comparative times spent on parts of the lessons in which the Drill and Experimental Groups were using word-for-word English stimuli and those when the stimuli were grammatical English sentences. Two possible differences in procedure might account for the time differences between Drill and Control Groups. One is

TABLE 2

Learning Time for Groups
(Mean Time in Minutes)

Group	Drill		Experimental		Control	
Material	A	B	A	B	A	B
Total Learning Time	230.74	234.71	186.78	198.26	196.77	217.11
Learning Time with Word-for-Word English	49.91	45.62	38.83	38.21	45.23	44.36
Learning Time with Grammatical English	72.63	76.34	58.06	64.22	56.44	62.48

the sheer mechanical difference in the machines used. The Answer-Mate, used by the control subjects, had the words one inch apart on cash register tape. A fast subject simply turns the knob as fast as possible and keeps up with the words as they come. The Min-Max II Teaching Machine proper, with whole sentences in each frame, also an inch apart, is turned only after the whole sentence has been translated, and thus does not pick up the momentum of continuous movement. In addition, the frames are typed on separate sheets and there is some delay every 10 frames between sheets. Another reason why the Drill Group might need more time has to do with the intellectual task. The subject may scan a series of words or a whole sentence before starting to translate. The delay caused by composing sentences should be greatest for those parts of the work where the English is presented in grammatical order, thus requiring revision of word order to produce the correct Russian. In the table the times are presented separately for parts of the material where the drill and experimental subjects had word-for-word order in stimuli and for the parts where they had grammatical English. Although controls take less time than drills throughout, the difference is significant only in the case where the drills were translating from grammatical English. The Experimental Group, however, is faster than the Drill Group throughout and the differences are significant at the 1% level. In the case of word-for-word English, the experimentals are also significantly faster than the controls. It is likely that the reason for this faster learning time for the experimental subjects is related to the fact that they always learned the two sets of words in entirely separate lists. The drill and control subjects were learning all words by the same method and therefore had them combined in "whole" lists, whereas the experimentals had only one of the two "part" lists at any one time and were never required to combine them. Thus the "time to combine", which requires additional trials in part-whole learning, is eliminated from their time scores.

Table 3 shows the analysis of variance for total learning time between Drill and Experimental Groups and Table 4 shows the analysis for those parts where the Drill and Experimental Groups were working with word-for-word (ungrammatical) English stimuli. Table 5 compares Drill and Experimental Groups where both were using grammatical English stimuli. This table also shows the same comparison of Drill and Control Groups. Since the drill subjects here take significantly longer than the controls at the 1% level of confidence, it seems safe to say that the intellectual problem of sentence translation makes at least part of the difference between the total learning time for the two groups. Although the other differences between these two groups are not significant, we cannot rule out the possibility that the mechanical differences also contribute.

Testing Procedure

All subjects were given the same series of tests in the last experimental session. The entire session was recorded on tape. The procedures and measures are described below in the order they were done.

Free Recall. At the beginning of the final session each subject was asked to recite as many Russian words as she could remember, in three minutes, in whatever order they came to mind. This provided some measure of vocabulary learning (number of words recalled) but also an opportunity to study the sequence of recall and measure the amount of "clustering" by form-class. Although all of our groups were expected to have some impression of form-classes from the translations, the groups with the pat-

TABLE 3

Analysis of Variance

Total Learning Time

Drill vs Experimental

Source	SS	df	MS	F	p
Method	12929.53	1	12929.53	9.69	$<.01$
Material	477.64	1	477.64	.36	ns
Interaction	112.90	1	112.90	.08	ns
Within Groups	37344.86	28	1333.74		
Total	50864.93	31			

TABLE 4

Analysis of Variance

Learning Time with Word-for-Word English

Drill vs Experimental

Source	SS	df	MS	F	p
Method	683.76	1	683.76	8.11	<.01
Material	48.21	1	48.21	.57	ns
Interaction	26.90	1	26.90	.32	ns
Within Groups	2360.87	28	84.32		
Total	3119.74	31			

Experimental vs Control

Source	SS	df	MS	F	p
Method	3507.45	1	3507.45	41.48	<.01
Material	58.70	1	58.70	.69	ns
Interaction	88.71	1	88.71	1.05	ns
Within Groups	2367.54	28	84.56		
Total	6022.40	31			

TABLE 5

Analysis of Variance

Learning Time with Grammatical English

Drill vs. Control

Source	SS	df	MS	F	p
Method	1806.60	1	1806.60	14.54	$<.01$
Material	190.03	1	190.03	1.53	ns
Interaction	10.87	1	10.87	.87	ns
Within Groups	3480.20	28	124.29		
Total	5487.70	31			

Drill vs Experimental

Source	SS	df	MS	F	p
Method	1425.24	1	1425.24	11.48	$<.01$
Material	194.93	1	194.93	1.57	ns
Interaction	12.06	1	12.06	.10	ns
Within Groups	3476.05	28	124.14		
Total	5108.28	31			

tern drills were expected to show greater clustering by form-class. There was no reason for predicting a difference in the number of words recalled, however.

English to Russian I. The basic test consisted of a series of 54 English sentences presented on the teaching machine to be translated into Russian. All the translations could be done with the three patterns to which the subject had been exposed. All required exclusively words from the "y" set. This meant that they were all combinations never before encountered by any group of subjects. No two of these words had ever before been used in the same sentence, so no sequential associations could have been formed. The subject was instructed to translate each sentence and proceed immediately to the next. This time no corrections were offered by the experimenter. The time to complete the series was as important as the accuracy because the basic prediction was that drill and experimental subjects would be able to produce Russian sentences much faster than control subjects. Several measures were derived from this test. Total time, number of sentences completely correct, sentences in which word order was incorrect, and number of individual words omitted or incorrectly translated, were included among the scores. Drill and experimental subjects were expected to do better with order but no difference between groups was predicted on vocabulary.

English to Russian II. A second translation task was presented after the first. This time instead of presenting English sentences which closely resembled those used in demonstrations and in the drill training, 18 English sentences were purposely composed so as to be as different as possible from those already seen yet still able to be translated by the familiar patterns of Russian sentences. Subjects were warned that the English would be unfamiliar: that, for example, instead of the sentence "You are not allowed to smoke," they would see "Hey you! Stop smoking." It was pointed out to them that these could both be roughly translated by the same familiar Russian sentence.

The purpose of this test was to determine whether the skill of the drill subjects consisted only of a quick transformation from one standard word order in English to the standard Russian pattern or whether the subject was in some sense generating a Russian sentence to express the meaning of the English sentence. Since the English words now commonly came in different order from the way they had been before,no mechanical transposition rule would be much help in Test II.

Cloze Test. Subjects were next shown on the machine a series of 22 Russian "sentences" each with one word omitted. They were instructed to read the sentences substituting an appropriate word in the blank. Again all the words used were from the "y" set and most of the sentence frames were in those sentence patterns which differed from English. The subject could, of course, supply "x" words for the blanks if she chose.

The scoring of this test was on the basis of form-class alone, since there was no point in designating any single word as correct. The 'correct' form-class was defined for most slots as being one of the six words used there in our materials. However, in each set of materials there were one or two pairs of classes which were equivalent as far as the translation could indicate and therefore as far as the Control Group could have known. These were treated as one form-class. Number of substitutions of appropriate words, that is, words of the correct form-class, was used as the accuracy score. Total time to complete the last 20 crucial items was also measured. It was predicted that the pattern drill subjects would make faster and more grammatically accurate substitutions than controls.

Memory Span. The memory span test could not take the conventional form of a gradually increasing number of words because in order to maintain grammatical patterns, only 3- and 4-word series could be used. In addition, 3-word sequences were easy enough to be done correctly by all subjects regardless of grammaticality. Therefore, the test consisted of sentences of the 4-word, unlike-English pattern, "y" words only. Two types of items were used: nine normal grammatical sentences and the same nine sentences presented backwards so as to be ungrammatical. The sequences were read to the subjects by the experimenter in list intonation at the rate of one word per second. The subject was instructed simply to repeat the words she heard in the same order she heard them.

The two scores were the numbers of correctly repeated sequences of forward and backward sets. If the drill subjects had acquired a "feeling" for word order more like their native language habits than that of the controls, they should have had higher scores on the forward sequences, but there would be no reason to expect a difference between the two groups on the backward sequences.

Russian to English Translation. Memory span has provided a partial test of ability to imitate sequences but it is not a test of understanding in the sense of comprehension of meaning. The most straightforward way to test this kind of comprehension is a translation from spoken Russian sentences into English. For this test the experimenter again read Russian sentences at the rate of one word per second. This time only forward sequences were used and they were read in sentence intonation. The subject was instructed to listen and give an English sentence that meant the same. After two practice sentences, 20 sentences of all three types were read. As soon as the subject finished one translation the next Russian sentence was read. In this way measures of both speed and accuracy could be obtained. Errors were scored separately according to whether they reflected vocabulary or grammatical difficulties. It was expected that pattern drills would produce better comprehension and therefore better and faster translations.

Results

Statistical Tests. A word is in order about the general policies followed in testing the results and in presenting them here. Most of the scores for the three method groups were subjected to 2x2 (method by material) analyses of variance comparing each of the three possible pairs of groups. The separate comparisons were made rather than the single 3x2 analysis for each measure because our interest was focused on the difference between each pair of methods rather than upon the trends across three groups. Bartlett's test for homogeneity of variance was applied beforehand and in those instances where the inhomogeneity was significant at the 1% level, we decided to compare the method and material groups separately using the Mann-Whitney U test. Whenever an analysis of variance showed a significant interaction of method and material which had interesting implications for our hypotheses, t tests were performed between the relevant sub-groups. All significance levels reported are based upon one tail tests. Detailed results of the analyses of variance are shown in tables only in those cases where results were significant at at least the 5% level.

Aptitude. The predominance of high scores both in totals and in separate parts I, II and V of the MLAT (see Figs. 2 & 3), where our subjects piled up near the maximum scores, produced considerable deviation from normality. This, together with the small number of subjects, led us to use rank order correlations to assess the relation between these scores and our various tests. Ties in rank will, of course, limit even these correlations in the case of those three parts of the test.

Table 6 shows only those rank order correlations with our various experimental measures which reached significance at at least the 5% level of confidence (by a 1 tailed test). Those underlined were significant at the 1% level. The total test was significantly correlated with 15 out of the 19 measures used. Part I, Number Learning, a kind of specialized vocabulary task, also significantly correlated with 15 and Part II, Phonetic Script, with 15 of our measures. It is interesting that Part II had one of the smaller distributions among our subjects, so this correlation was in spite of a narrow range of talent on MLAT. Part III, Spelling Clues, was correlated with 11, Part IV, Words in Sentences with 5 and Part V, Paired Associates, with 7. Positive correlations in all cases indicate that better scores on the MLAT go with better scores on our measures regardless of the direction of actual numerical values.

Tables 6a, 6b and 6c show rank order correlations separately for Drill, Experimental and Control Groups respectively. On the total test the correlation was greater for the Control than for the Drill Group with 13 out of 19 measures. Eight of the 19 correlations for the Control Group were significant at the 1% level and 5 more at the 5% level. (One was significantly negative at the 5% level.) In the case of the Drill Group none was significant at the 1% level and only 5 at the 5% level. The Experimental Group stood between with 2 at the 1% level and 12 at the 5% level.

Free Recall. The first row in Table 7 shows the mean number of words recalled by drill, experimental and control subjects learning A and B materials. The analysis of variance showed no significant differences among the groups. The second row of the table shows the mean number of repetitions in each subject's sequence of recall. A "repetition" is an instance where a word of a given form-class follows in the sequence a word of that same class. The more repetitions, the greater the degree of "clustering". For each subject the number of repetitions to be expected by chance, given the particular number of words she recalled in each category, was also calculated. The third row in the table shows the mean expected repetitions for each group. A t test shows that the actual number of repetitions was, on the average, significantly higher at the 1% level than the expected number. Our subjects did tend to cluster words by form-class in free recall. However, analyses of variance, shown in Table 8 on the difference between expected and observed number of repetitions, showed that the only significant effect of method was greater clustering for drill than for experimental subjects. There were also two interaction effects between material and method in the comparison of the Experimental Group with each of the other two.

English to Russian I. Table 9 shows several types of scores based on the subjects' performance on 54 sentences to be translated at sight from English to Russian. The first row shows the mean time to complete the series for the six groups. Table 10 shows results of analyses of variance of these scores. Comparing Drill with Control Groups the variance contributed by the difference in training methods was significant at the 1% level. In addition, the shorter time for the B set of materials was significant at

TABLE 6

Aptitude Correlations
All Ss All Parts

MLAT	Total	I	II	III	IV	V
Test						
Learning Time	.61	.44	.42	.50		.35
E to R Part I						
Time	.47		.24	.39	.30	.26
Correct		.27				.26
Order Errors						
Vocabulary Errors	.58	.42	.37	.38	.32	.32
E to R Part II						
Time	.49	.25	.26	.44	.26	
Correct	.41	.48	.51		.24	
Order Errors		.25				
Vocabulary Errors	.57	.47	.52	.26	.39	.24
R to E						
Time	.41		.26	.33		
Correct	.39	.37	.52			
Grammatical Errors	.34	.30	.44	.32		
Vocabulary Errors	.43	.40	.48			.34
Memory Span						
Total	.34	.29	.33	.28		
Forward	.52	.38	.47	.39		
Backward						
Cloze Test						
Time	.58	.43	.48	.45		
Correct	.46	.43	.40			.25
Free Recall						
Total Words	.31	.28	.31	.31		

TABLE 6a

Aptitude Correlations
Drill Ss All Parts

MLAT	Total	I	II	III	IV	V
Test						
Learning Time						
E to R Part I Time						
Correct				.43		
Order Errors				.49		
Vocabulary Errors						
E to R Part II Time	.55		.47			
Correct		.58	.52			
Order Errors						
Vocabulary Errors	.54	.69	.63			
R to E Time						
Correct			.44			
Grammatical Errors	.47	.52		.49		
Vocabulary Errors			.50			
Memory Span Total						
Forward		.51				
Backward						
Cloze Test Time	.49	.64	.46			
Correct	.49	.48		.47		
Free Recall Total Words		.47				

TABLE 6b

Aptitude Correlations
Experimental Ss All Parts

MLAT	Total	I	II	III	IV	V
Test						
Learning Time	.51					
E to R Part I						
Time	.56				.49	
Correct	.58	.52				
Order Errors						
Vocabulary Errors	.70	.53			.51	
E to R Part II						
Time	.46					
Correct	.57	.57	.65			
Order Errors	.51	.56	.63			
Vocabulary Errors	.52				.49	
R to E						
Time	.52					
Correct	.60	.51	.59			
Grammatical Errors						
Vocabulary Errors	.54		.46			.43
Memory Span						
Total						
Forward	.45			.57		
Backward						
Cloze Test						
Time	.54		.66			
Correct	.49		.55			
Free Recall						
Total Words						

TABLE 6c

Aptitude Correlations
Control Ss All Parts

MLAT	Total	I	II	III	IV	V
Test						
Learning Time	.88	.76	.66	.77		.55
E to R Part I						
Time	.54			.55	.43	
Correct						
Order Errors	—.43					
Vocabulary Errors	.49					
E to R Part II						
Time	.46			.72		
Correct						
Order Errors						
Vocabulary Errors	.71	.55	.60	.44		.54
R to E						
Time	.61			.52		
Correct	.55	.53	.58			
Grammatical Errors	.42		.65			
Vocabulary Errors	.64	.75	.56			.65
Memory Span						
Total	.66	.47	.54	.63		
Forward	.74	.58	.59	.72		
Backward	.66	.46	.48	.53		
Cloze Test						
Time	.74	.72		.74		.50
Correct		.55				.55
Free Recall						
Total Words			.48			

TABLE 7

Free Recall

Group	Drill		Experimental		Control	
Material	A	B	A	B	A	B
Mean Number of Words Recalled	42.88	46.25	45.25	42.25	41.12	42.38
Mean Number of Repetitions	17.88	20.62	17.25	11.88	15.50	16.62
Mean Number of Expected Repetitions	4.52	4.36	5.08	3.89	4.54	4.22
R - E	13.36	16.26	12.17	7.99	10.96	12.40

TABLE 8

Analysis of Variance

Clustering in Free Recall

$\Sigma r - E(R)$

Drill vs Experimental

Source	SS	df	MS	F	p
Method	178.61	1	178.61	8.22	<.01
Material	3.25	1	3.25	.15	ns
Interaction	100.82	1	100.82	4.64	<.05
Within Groups	608.10	28	21.72		
Total	890.78	31			

Experimental vs Control

Source	SS	df	MS	F	p
Method	20.48	1	20.48	1.57	ns
Material	15.13	1	15.13	1.16	ns
Interaction	63.28	1	63.28	4.86	<.05
Within Groups	364.40	28	13.01		
Total	463.29	31			

TABLE 9

English to Russian I

Group	Drill		Experimental		Control	
Material	A	B	A	B	A	B
Mean Time in Minutes	6.98	6.94	9.07	9.22	11.80	9.35
Mean Number of Sentences Correct	48.62	48.38	35.75	39.12	18.50	25.25
Mean Number of Order Errors	.88	.38	13.00	2.38	27.50	18.25
Mean Number of Vocabulary Errors	5.00	6.75	10.00	17.25	21.25	18.75

TABLE 10

Analysis of Variance

English to Russian I - Time

Drill vs Control

Source	SS	df	MS	F	p
Method	104.83	1	104.83	37.04	$< .01$
Material	12.40	1	12.40	4.38	$< .05$
Interaction	11.57	1	11.57	4.09	ns
Within Groups	79.31	28	2.83		
Total	208.11	31			

Drill vs Experimental

Source	SS	df	MS	F	p
Method	38.26	1	38.26	11.59	$< .01$
Material	.03	1	.03	.01	ns
Interaction	.08	1	.08	.02	ns
Within Groups	92.36	28	3.30		
Total	130.73	31			

the 5% level, with most of the difference occurring in the Control Group. However, the interaction was not significant. The Experimental Group stood between the other two, significantly different from the Drill Group at the 1% level, but not significantly different from the controls.

The second row of Table 9 shows the mean number of sentences correct out of 54. This time there was again a significant effect of method (see Table 11) reaching the 1% level for all three intergroup comparisons with the Experimental Group again standing between the other two in total sentences correct. The effect of materials was insignificant.

The third row of Table 9 shows the mean number of errors in word order for the six groups. The very small number of such errors for the Drill Groups compared with the Control Groups is obvious. Only 4 of the 16 drill subjects made any at all. This means, however, that the variance in the Drill Group was severely limited and the variance in the six groups could not be considered homogeneous. Therefore, instead of an analysis of variance, the Drill and Control Groups and the A and B Groups were compared separately by means of a Mann-Whitney U test. The difference between Drill and Control Groups was significant at the 1% level, whereas the difference between the A and B sets of materials was non-significant. The Mann-Whitney U test was also used to compare the drill with the experimental subjects and the experimental with the control. Again in both cases the differences between the methods were significant at the 1% level and between materials non-significant.

The last row of Table 9 shows the number of purely vocabulary errors. Again the heterogeneity of variance between drills and controls made analysis of variance questionable, and the Mann-Whitney U test was used instead. The Drill Group had significantly fewer errors than the Control at the 1% level and no difference was found between the two sets of materials. In the case of vocabulary, analyses of variance showed that the difference between the drill and experimental subjects was significant at the 5% level (see Table 12) but between experimental and control there was no significant difference.

English to Russian II. Table 13 shows the results of the translation test in which various and unfamiliar forms of English sentences were used. Since the test was much shorter, only 18 sentences, the figures are not comparable in absolute value with those of Test I though they measure about the same skills. Again, the time in minutes to complete the series significantly favors the Drill Group over Control with 1% confidence. The difference between A and B materials was not significant. The analysis of variance is shown in Table 14. The Experimental Group stood between the other two groups but did not differ significantly from either in total time for Test II.

The analyses of variance shown in Table 15 on the number of sentences completely correct show a difference between Drill and Control Groups which was significant at the 1% level, but no significant difference between A and B materials. The Experimental Group also had significantly more sentences correct than the Control Group but significantly fewer than the Drill, both at the 1% level.

The errors of word order show inhomogeneity of variance significant at the 1% level. Mann-Whitney U tests comparing group differences showed all three method groups having significantly different numbers of order errors.

TABLE 11

Analysis of Variance

English to Russian I - Correct

Drill vs Control

Source	SS	df	MS	F	p
Method	5671.10	1	5671.10	106.78	$< .01$
Material	84.50	1	84.50	1.59	ns
Interaction	98.10	1	98.10	1.85	ns
Within Groups	1487.20	28	53.11		
Total	7340.90	31			

Drill vs Experimental

Source	SS	df	MS	F	p
Method	979.10	1	979.10	19.30	$< .01$
Material	19.60	1	19.60	.39	ns
Interaction	26.20	1	26.20	.52	ns
Within Groups	1420.10	28	50.72		
Total	2445.00	31			

Experimental vs Control

Source	SS	df	MS	F	p
Method	1937.50	1	1937.50	28.02	$< .01$
Material	205.00	1	205.00	2.96	ns
Interaction	22.80	1	22.80	.33	ns
Within Groups	1935.90	28	69.14		
Total	4101.20	31			

TABLE 12

Analysis of Variance

English to Russian I - Vocabulary Errors

Drill vs Experimental

Source	SS	df	MS	F	p
Method	480.50	1	480.50	7.42	$<.05$
Material	162.00	1	162.00	2.50	ns
Interaction	60.50	1	60.50	.93	ns
Within Groups	1813.00	28	64.75		
Total	2516.00	31			

TABLE 13

English to Russian II

Group	Drill		Experimental		Control	
Material	A	B	A	B	A	B
Mean Time in Minutes	3.77	4.10	4.30	4.12	5.01	4.45
Mean Number of Sentences Correct	13.75	12.25	8.88	9.75	6.00	4.88
Mean Number of Order Errors	0	1.00	4.62	2.25	7.62	7.25
Mean Number of Vocabulary Errors	4.75	6.75	9.12	10.50	12.38	11.25

TABLE 14

Analysis of Variance

English to Russian II - Time

Drill vs Control

Source	SS	df	MS	F	p
Method	5.07	1	5.07	8.74	$<.01$
Material	.11	1	.11	.19	ns
Interaction	1.57	1	1.57	2.71	ns
Within Groups	16.27	28	.58		
Total	23.02	31			

TABLE 15

Analysis of Variance

English to Russian II - Correct

Drill vs Control

Source	SS	df	MS	F	p
Method	457.60	1	457.60	84.43	<.01
Material	13.80	1	13.80	2.55	ns
Interaction	.20	1	.20	.04	ns
Within Groups	151.90	28	5.42		
Total	623.50	31			

Drill vs Experimental

Source	SS	df	MS	F	p
Method	108.80	1	108.80	10.11	<.01
Material	.80	1	.80	.07	ns
Interaction	11.20	1	11.20	1.04	ns
Within Groups	301.40	28	10.76		
Total	422.20	31			

Experimental vs Control

Source	SS	df	MS	F	p
Method	120.10	1	120.10	11.71	<.01
Material	.10	1	.10	.01	ns
Interaction	8.10	1	8.10	.79	ns
Within Groups	287.20	28	10.26		
Total	415.50	31			

The vocabulary errors were compared by an analysis of variance shown in Table 16. The difference between the Drill and Control Groups was significant at the 1% level, but the materials made no significant difference. The Experimental Group had significantly more vocabulary errors than the Drill Group (at the 5% level) but not significantly fewer than the Control.

Cloze Test. In the Cloze test, where subjects chose appropriate words to fill blanks in the three kinds of sentences, two scores were obtained. One was time to complete the set of 20 items, and the other was number of "correct" or grammatically appropriate words substituted. Table 17 shows these data for the six groups. The Drill Group took less time and had more correct substitutions than the Control Group. The Experimental Group did not take significantly longer than the Drill Group but did significantly surpass the Control Group at the 1% level. However, when it came to total number of grammatically correct words inserted, the experimental subjects did less well than the drill subjects with 5% significance and insignificantly better than the controls. Tables 18 and 19 show the analyses of variance. Both differences between groups were significant at the 1% level. Materials made no significant difference except in the comparison of the experimental and control subjects, where the B materials were completed significantly faster (5% level). In the case of the drill-control comparison there was an interaction effect on number correct between materials and method which was significant at the 5% level.

In the Cloze test subjects were free to select any Russian words that came to mind and were therefore not confined to the "y" words as they were in other tests. Since they had had much more practice with the "x" words because of the requirements of the training procedure, it might have been expected that they would rely primarily on the "x" words to fill the blanks. Table 20 shows that this was the reverse of what happened. The strong preference for "y" words was significant for all three groups using a t test of the individual x-y differences. It seems probable that the reason for this is that the Cloze test came just after exposure to the English to Russian translation tests involving exclusively "y" words. Thus within the test session many more "y" words had been already produced by the subject. In addition, of course, the Cloze test sentences were made up of "y" words only. A confirmation of the effects of the test session is found in comparing the proportions of "x" and "y" words used in the Cloze test with the proportions that had appeared in free recall. Since the free recall test came first in the test session, there is no reason to attribute the distribution of responses to anything but relative strengths of habits from past sessions. Although the "x" set had had much more previous exposure, there was no apparent difference in free recall favoring either set of words. Subjects in all groups recalled roughly two thirds of both sets of words. No differences between numbers of "x" and "y" words would be significant. It seems that the shift toward choosing "y" words in the Cloze test was due to recall from earlier tests in the series.

Memory Span. The data on memory span represent simply the number of 4-word sequences that were correctly repeated. Since no variation was possible in the number of words in grammatical sequences, a standard memory span measure was not possible. All the sequences were chosen from that set of 4-word sentences which had a word order different from English. They were presented in forward order for grammatical sequence and in backward order for ungrammatical sequence. Each particular set of words

TABLE 16

Analysis of Variance

English to Russian II - Vocabulary Errors

Drill vs Control

Source	SS	df	MS	F	p
Method	294.10	1	294.10	13.72	<.01
Material	1.60	1	1.60	.07	ns
Interaction	19.40	1	19.40	.90	ns
Within Groups	600.40	28	21.44		
Total	915.50	31			

Drill vs Experimental

Source	SS	df	MS	F	p
Method	132.10	1	132.10	7.49	<.05
Material	22.80	1	22.80	1.29	ns
Interaction	.70	1	.70	.04	ns
Within Groups	493.90	28	17.64		
Total	649.50	31			

TABLE 17

Cloze Test

Group	Drill		Experimental		Control	
Material	A	B	A	B	A	B
Mean Time in Min.	3.36	3.57	3.83	3.51	4.74	4.03
Mean No. Correct	18.00	17.75	15.12	15.00	12.88	14.88

TABLE 18

Analysis of Variance

Cloze Test - Time

Drill vs Control

Source	SS	df	MS	F	p
Method	6.81	1	6.81	21.97	$<.01$
Materials	.50	1	.50	1.61	ns
Interaction	1.72	1	1.72	5.55	$<.05$
Within Groups	8.63	28	.31		
Total	17.66	31			

Experimental vs Control

Source	SS	df	MS	F	p
Method	4.11	1	4.11	9.34	$<.01$
Materials	2.13	1	2.13	4.84	$<.05$
Interaction	.32	1	.32	.73	ns
Within Groups	12.34	28	.44		
Total	18.90	31			

TABLE 19

Analysis of Variance

Cloze Test - Correct

Drill vs Control

Source	SS	df	MS	F	p
Method	128.00	1	128.00	17.46	<.01
Material	6.10	1	6.10	.83	ns
Interaction	10.20	1	10.20	1.39	ns
Within Groups	205.20	28	7.33		
Total	349.50	31			

Drill vs Experimental

Source	SS	df	MS	F	p
Method	63.30	1	63.30	7.14	< .05
Material	.30	1	.30	.03	ns
Interaction	.00	1	.00	.00	ns
Within Groups	248.40	28	8.87		
Total	312.00	31			

TABLE 20

Mean Numbers of "x" and "y" Words Selected by Subjects in Tests Where a Choice Was Permitted

	Drill	Experimental	Control
Free Recall			
"x" words	22.69	21.19	20.88
"y" words	21.88	22.56	20.88
Cloze Test			
"x" words	7.44	7.06	5.32
"y" words	11.81	10.62	11.50

occurred in both orders. Table 21 shows the results. The analysis of variance for forward order is shown in Table 22. In backward, or (ungrammatical) order the drill subjects did slightly better than experimentals, who did better than controls, but none of the differences was significant. In forward, (or grammatical) order the Drill Group was superior to the controls (at the 1% level) and the Experimental Group was also better than the controls (at the 5% level) and insignificantly lower than the drills. Taking as the score for each individual subject the difference between forward and backward performance, the only significant difference (5% level) was between drills and controls (see Table 23). In none of these comparisons was there any significant difference between the two sets of material.

Spoken Russian to English. The results comparing scores for translation into English from spoken Russian sentences showed relatively few differences among the groups. Table 24 shows the means of four measures. As for the time to complete the sentences, neither the method nor the materials showed any very considerable difference in any group comparison. Table 25 shows the analysis of variance for Drill and Control Groups. Neither method nor materials produced a significant effect, but for some reason the interaction was significant at the 1% level. The A materials were done faster by the Drill Group and the B materials by the Control Group. In the case of the Control Group the difference between the two materials was significant at the 5% level (t test) but for the Drill Group it was not significant.

The number of sentences completely correct was in decreasing order for Drill, Experimental and Control Groups and was greater for the B materials, but none of these differences reached significance.

Grammatical errors were separately tabulated. These included a variety of errors not attributable to particular words: errors of tenses of verbs, misinterpretations of negative sentences and so forth. They did not correspond exactly to order errors in the English to Russian translations because they were not all attributable to the order of Russian words, nor did they consist of giving English sentences in incorrect order. Heterogeneity of variance made it necessary to use the Mann-Whitney U test to analyze the significance of these data. The method of training made no significant difference. However, there were significantly more grammatical errors on the A materials with a probability of less than .01.

Vocabulary errors were assessed by analysis of variance, shown in Table 26. Here the method did make a difference, with the Drill Group having fewer errors than the Control Group, significant at the 5% level. The other group differences were not significant.

Comparison of Patterns. All the analyses so far presented have been based upon all the sentence patterns used in the test. However, the three patterns differed along two dimensions and it would be plausible to predict differences in their results. The 3-word pattern could be expected to be easier than the 4-word pattern when both had a grammatical order unlike English. The 4-word pattern which had a word order corresponding to correct English might be easier than the 4-word pattern that was unlike English. Table 27 shows the results of a number of the tests separately for the different sentence patterns. Analyses of variance performed separately for the three types of sentences showed that the effect of method (in all but 2 out of 27 comparisons) was significant at at least the 5% level for all three types of sentences on time and total sentences correct in English to Russian Translation Part I and on total correct for the English to Russian

TABLE 21

Memory Span - Sequences Correct
(Mean Number Correct)

Group	Drill		Experimental		Control	
Material	A	B	A	B	A	B
Forward	7.62	6.88	7.50	5.88	4.62	5.38
Backward	5.88	6.12	6.38	5.25	4.88	5.00

TABLE 22

Analysis of Variance

Memory Span - Forward

Drill vs Control

Source	SS	df	MS	F	p
Method	40.50	1	40.50	10.44	$< .01$
Material	0.00	1	0.00	0.00	ns
Interaction	4.50	1	4.50	1.16	ns
Within Groups	108.50	28	3.88		
Total	153.50	31			

Experimental vs Control

Source	SS	df	MS	F	p
Method	22.80	1	22.80	6.61	$< .05$
Material	1.50	1	1.50	.43	ns
Interaction	11.30	1	11.30	3.28	ns
Within Groups	96.60	28	3.45		
Total	132.20	31			

TABLE 23

Analysis of Variance

Memory Span - Forward Correct Less Backward Correct

Drill vs Control

Source	SS	df	MS	F	p
Method	11.30	1	11.30	5.23	$< .05$
Material	.30	1	.30	.14	ns
Interaction	5.20	1	5.20	2.41	ns
Within Groups	60.40	28	2.16		
Total	77.20	31			

TABLE 24

Russian to English
(Means-Time in Minutes)

Group	Drill		Experimental		Control	
Material	A	B	A	B	A	B
Time	4.12	4.71	4.32	4.16	4.94	4.15
Correct	10.12	12.12	9.62	10.88	7.00	10.62
Grammatical Errors	5.75	.88	4.12	1.00	6.50	2.12
Vocabulary Errors	9.50	10.38	12.12	13.50	17.38	14.25

TABLE 25

Analysis of Variance

Russian to English - Time

Drill vs Control

Source	SS	df	MS	F	p
Method	.13	1	.13	.27	ns
Material	.08	1	.08	.16	ns
Interaction	3.76	1	3.76	7.67	<.01
Within Groups	13.78	28	.49		
Total	17.75	31			

TABLE 26

Analysis of Variance

Russian to English - Vocabulary

Drill vs. Control

Source	SS	df	MS	F	p
Method	276.10	1	276.10	5.12	$< .05$
Material	10.10	1	10.10	.19	ns
Interaction	32.10	1	32.10	.70	ns
Within Groups	1509.20	28	53.90		
Total	1827.50	31			

TABLE 27

Pattern Comparisons

	Test	Method	3-Word Unlike English	4-Word Like English	4-Word Unlike English
English to Russian Part I					
	Time	Drill	1.77	2.13	2.61
		Experimental	2.59	2.76	3.25
		Control	2.88	3.24	3.91
	Sentences Correct	Drill	16.68	16.62	15.19
		Experimental	13.19	13.38	10.88
		Control	7.62	8.69	5.56
	Order Errors	Drill	0.00	0.00	.62
		Experimental	1.69	1.19	4.81
		Control	7.44	6.44	9.00
	Vocabulary Errors	Drill	1.69	1.38	2.81
		Experimental	4.25	4.44	4.94
		Control	4.56	6.06	9.38
English to Russian Part II					
	Sentences Correct	Drill	4.56	4.31	4.12
		Experimental	3.56	3.00	2.75
		Control	2.31	2.00	1.12
	Order Errors	Drill	.00	.38	.12
		Experimental	.31	.81	2.31
		Control	2.06	2.44	2.94
	Vocabulary Errors	Drill	1.62	1.81	2.31
		Experimental	3.12	3.50	3.19
		Control	2.38	4.00	5.44
Russian to English					
	Sentences Correct	Drill	6.31		3.69
		Experimental	6.00		3.44
		Control	5.19		2.38
	Grammatical Errors	Drill	.94		2.38
		Experimental	.62		1.75
		Control	1.62		2.38
	Vocabulary Errors	Drill	2.88		6.12
		Experimental	3.31		7.75
		Control	3.94		10.94

Part II. Scattered significant differences also occurred on other tests. However, when the three method groups are compared with one another on scores representing the differences between performance on 3-and 4-word sentences or on scores representing the differences between English order and non-English order, there is little indication that length or order had more effect on one group than another. Although we might have expected that the control subjects would have been at a greater disadvantage on the more difficult sentences, especially where the difficulty was a matter of word order, there is no indication that this was so.

Discussion

Aptitude. In spite of the fact that our experimental measures were extremely limited in the amount of material covered and in the variety of language skills tapped, they showed quite respectable correlations with the whole and with some parts of the MLAT. These results encouraged us to think that we were measuring skills which required some of the same abilities as studying foreign languages.

The differences in aptitude correlations between our method groups may also be of interest. Poltizer (personal communication) has expressed the view that pattern drills will have a leveling effect upon language learners with respect to aptitude, that is, the less apt will succeed much better with drill techniques than without while those with greater original aptitude get less benefit from directed drills. If this is true, then aptitude might be expected to correlate with our measures more closely for control than for pattern drill subjects since the leveling effect of the drill would reduce the range of scores for drill subjects. The correlations show just such a trend with higher and more significant ones for the Control

Group than for the Drill Group with the Experimental Group standing between. These results are consistent with the hypothesis that ability is a smaller factor in the success of the drill method and also imply that differences within our Experimental Group may be more dependent on ability than those in the Drill Group.

The one part of the MLAT, however, that might be interpreted most readily as a test of grammatical ability, Part IV, Words in Sentences, showed almost no relation to any of our measures, even "order errors" and "grammatical errors", which were specifically tests of grammatical skill in Russian. Although it might be expected that one reason for this lack of correlation would be the very small variance of order errors in the Drill Group, the separate correlations for the Control Group were also nearly all insignificant. As far as our measures go, therefore, we find no good evidence that this, or, in fact, any other _part_ of the MLAT is related to our subjects' ability to learn to put Russian words in Russian order.

Free Recall: Our free recall test yielded no scores that gave any support to our hypotheses. The lack of difference shows that we succeeded in keeping at least this aspect of vocabulary learning under control although some vocabulary differences showed up on other tests. Also the fact that "x" and "y" words were recalled about equally often suggests that we succeeded in bringing both sets to a similar level of learning. However, our hypothesis that subjects learning by pattern drills would have a better sense of the form-class of individual words found no support in the number of form-class "repetitions" in recall sequence. The only significant simple effect of method was a greater clustering ratio for drill subjects than for experimentals. It is difficult to interpret this in terms of our hypothesis.

Although the drill group had more pattern drills than the experimentals, the controls had none at all. The most likely explanation of the lower clustering ratio for experimentals seems to lie in the fact that each set of words from one form-class was divided into two smaller sets in the learning procedure and separated both in time and in method of training. Interactions between material and method in comparisons between experimentals and the other groups also seem to have no direct implications for our hypotheses. All groups showed a significant degree of clustering by form-class, as though each had acquired some sense of these categories. The English translations of the words and possibly the grammar lessons must have provided enough cues to account for this effect without pattern drill.

English to Russian. Our hypotheses stated that the Drill and Experimental Groups, having acquired language habits more like those of the native language, should be able to produce sentences faster and with fewer errors in word order than the controls. Although control subjects had demonstrated that they were able to apply the rules and had equal exposure to the words, it was expected that this knowledge would not be enough for them to equal the fluency in sentences of the other subjects. Control subjects should know vocabulary as well, but their native language habits of word order could be expected to mislead them in constructing sentences when Russian grammar differs from English.

The results support the hypotheses as far as time and order errors are concerned for Drill and Control Groups. On Part I the drill subjects took about two-thirds as much time as controls and averaged less than 6 sentences out of 54 with any kind of errors as compared with 32 for controls. The Experimental Group, however, did significantly less well than the Drill

Group on both measures. Their time was very close to that of the controls and insignificantly different. In total correct they stand roughly midway between the other groups and significantly different from both at the 1% level of confidence. The difference between the Drill and Experimental Groups was not predicted in the hypothesis. It had been supposed that the superiority of pattern drills would depend primarily on the fact that the patterns were drilled rather than upon the way in which the test words were learned. Although it is true that the Drill Group had more actual pattern practice, the Experimental Group had had enough such practice with the "x" words to reach a plateau of speed and accuracy comparable with that of the Drill Group in the training period. It seems at least equally plausible to attribute the difference to the method of learning the "y" or test words. We will return to this question in the following discussion of vocabulary errors.

The analysis of vocabulary errors, that is, simply missing words or selecting the wrong word, showed a large and significant difference favoring the Drill Group over the Control. The Drill Group also had significantly fewer errors than the Experimental Group (at the 5% level), but the Experimental Group was not significantly better than the Control. Similar results occurred also when variations on the English sentences were used. Since the number of exposures to each individual word was equal for the three groups, we must explain the poorer performance of the experimentals and controls in some way. At least two different explanations seem possible, one in terms of the conditions of learning, the other in terms of the test.

First, it is possible that words are more readily learned in sentence context. New words in the native language are more often encountered in sentences. Here the normal cues of form-class are provided instead of only the artificial cues of translations into a foreign language. The sentence context may also add to the meaning. If ease of learning in sentence context is the explanation of the difference between vocabulary performance in our three groups, the same difference should be found if we look at error scores during the learning sessions. Table 28 compares some vocabulary learning scores with the data on test errors. The first row shows the number of vocabulary errors in the very first learning sessions with 48 of the words. Although the Control Group made more errors than either Drill or Experimental, analysis of variance shows neither method nor materials to produce a significant difference in errors. The second row shows errors in the final learning session for the same words used in the test. Again controls and experimentals make more errors but neither method nor material has a significant effect. The last row shows the mean differences between each subject's vocabulary errors on the test and the same subject's errors in the last learning session. This is a rough measure of forgetting during that 24 hour period. Analysis of variance between Drill and Control Groups was not possible because of significant heterogeneity of variance, but a Mann-Whitney U test shows the loss for the Control Group to be significantly greater at the 1% level. Analyses of variance were used to compare the Experimental Group with the other two. Although the figures show loss about midway between those of the Drill and Control Groups, the difference is significant only in comparison with the Drill Group (1% level, Table 29). This implies that the superiority of the Drill Group in vocabulary is not entirely due to a greater ease of vocabulary learning under drill conditions but is, in part at least, a difference in retention or transfer. It should be remembered, however, that there was no difference in word retention as measured by free recall.

Another possible explanation of experimental and control subjects' inferiority in vocabulary lies in the fact that the words are tested in a different context from that in which they were learned. In particular, the test requires the control subject to do two things at once, first, remember the words, something she has repeatedly demonstrated her ability to do during learning, and second, to think about the order in which they must come in the sentence. Although she has demonstrated that she can perform this task also, to do the two tasks at the same time introduces a new element of competition for both. For the drill subjects, in contrast, the test repeats almost exactly the training procedure except for particular word sequences and lack of feedback. These subjects had always had to work simultaneously on vocabulary and word order problems throughout training. If the drill method produced the kind of automatic, unconscious grammatical responses it was supposed to, the subjects would be able to give a greater share of their attention to the vocabulary task and succeed better even though their original learning of vocabulary was not significantly superior. The fact that in spite of their pattern drill training the experimental subjects were not able to do better than the controls as far as vocabulary is concerned suggests that vocabulary performance is not helped by pattern drills except when the words tested actually occur in the drills.

Cloze Test. Drill subjects showed the predicted superiority over controls in their ability to fill blanks in sentences with grammatically appropriate words. In this test the Experimental Group, although not significantly more accurate, was significantly faster than the Control Group

TABLE 28

Vocabulary Errors in Learning

Group	Drill		Experimental		Control	
Material	A	B	A	B	A	B
Errors in Early Learning "x" Words	25.75	15.50	24.25	22.50	32.38	34.50
Errors in Last Learning Session "y" Words	6.25	4.50	3.88	8.75	6.50	9.50
Errors in Test "y" Word Minus Errors in Last Learning Session	-1.25	2.25	6.12	8.50	14.75	9.25

TABLE 29

Analysis of Variance
Vocabulary Errors in Test: "y" Words Minus
Errors in Last Learning Session

Drill vs. Experimental

Source	SS	df	MS	F	p
Method	371.30	1	371.30	9.37	$<.01$
Material	69.00	1	69.00	1.74	ns
Interaction	2.50	1	2.50	.06	ns
Within Groups	1109.90	28	39.64		
Total	1552.70	31			

at the 1% level and not significantly below the Drill Group. Since subjects were instructed to fill the blanks with any appropriate word that came to mind, this test could be regarded as a kind of controlled recall test. Superiority in the Cloze Test might have been attributed to a greater available pool of words in memory. However, since there was no difference in free recall between the groups, it seems reasonable to attribute the Cloze Test differences to the better sense of form-class which we predicted the pattern drill subjects would acquire through repeated use of words in sentences.

Memory Span. Our hypothesis stated that the Drill and Experimental Groups, with stronger grammatical habits, would perceive material in larger units, or "chunks", and hence have a longer span of memory. The expectation would apply only to grammatical sequences however. The results were consistent with this hypothesis in that there were significantly more sequences correct for the Drill and Experimental than for the Control Group on forward but not on backward lists. (Drill and experimental were not significantly different on either forward or backward lists.) The fact that the pattern drill subjects did better than the controls on grammatical sequences but not on ungrammatical ones shows that their superiority cannot be attributed to a difference in vocabulary skill or to greater practice with remembering several words at a time. The difference must be attributed to practice with the particular grammatical patterns used in the test. It is worthwhile to repeat here that no particular sequences of words used in the test had ever been practiced, since only "y" words were used in the test. The practiced sequences were with different words of the same form-classes as those used in the test. The fact that the Experimental Group did better than the controls shows that there was some transfer from the pattern drills with "x" words to memory span with "y" words. The learning of grammatical structures seems to have some effect beyond the actual pool of words of which they were made up in training.

Spoken Russian to English. The translation from spoken Russian to English was designed as a test of comprehension. The subject had to perceive correctly the Russian sentences, as he did in memory span, but had also to understand their meaning. It was a difficult task, as any novice linguist can testify after his first exposure to native speakers in conversation. Our drill subjects had slightly better scores than the controls in this test, but only in vocabulary errors did the difference due to method reach significance at the 5% level. We do not, therefore, find evidence in this test that grammatical habits acquired by our drill subjects made them able to perceive the spoken language more meaningfully. The number of grammatical errors in Russian to English translation was significantly different (by the Mann-Whitney U test) for our two sets of materials. A large number of these errors consisted of leaving out negatives and past tense markers in the English. Since these happen to occur several times in the A set of materials and not at all in the B set, we can probably safely conclude that these errors account for the difference between the materials. Also, drill subjects were faster than controls on B materials and control subjects on A materials. This interaction is significant at the 1% level. These differences due to material, however, have no particular implications for our hypotheses.

Conclusions

Conclusions from the viewpoint of the linguist: The present experiment is based upon a very simple "theory" of grammar in which a sentence is defined as a sequence of "slots" into which words of given form-classes may be put. No more detailed analysis, either of phrase structure or of deep transformational structure, is implied. The corpus is much too small for such an analysis to have much meaning. The question therefore is whether such a limited treatment of grammar could have any implications for more extensive learning of the grammatical structure of a language.

Two arguments may be made in order to justify such a study. The first is that a beginning learner of a second language probably is forced to start with imitations of particular form-class sequences which are presented to him ready-made. He has no other basis for forming utterances. Secondly, it seems likely that a corpus of sentence-frames learned in this way would be necessary before any of the more fundamental relationships could be learned. How a knowledge of the transformational relationships in a given language could or should be used in designing lessons for the non-native speaker of that language is an interesting question admittedly much beyond the scale of the present study. Our argument is that an adult learner of a new language must work for a time, perhaps for a considerable time, with a grammar that is based upon a limited number of pattern sentences and is learned by imitation. Therefore we believe it is useful to the language teacher to know which methods of learning these initial patterns are most effective.

Politzer (1965) refers to Palmer's (1917) distinction between "primary matter", units learned by heart, and "secondary matter", units built or derived from primary matter. The present experiment deals only with the learning of primary matter and with one fairly simple secondary process, substitution. It does not suggest any other means of deriving secondary matter, but its assumptions do not require that substitution be regarded as the only possible process for deriving it.

The circumstances of second-language learning in the early stages make of it a very special case of language learning and language skill. Although presumably the very advanced speaker of a second language may approach the kind of grammatical skill the native speaker has, and although in his later study he may learn to apprehend the more complex and deeper levels of grammatical relation, it seems doubtful that the first steps of primary learning bear much resemblance to the later stages of secondary derivation. Therefore the learning processes of our subjects may be interesting to language teachers without necessarily having any important implications for comprehensive theories of grammar.

Conclusions from the viewpoint of the language teacher: Our results have several practical implications. The most important of these is a demonstration that something is accomplished by a pattern drill. Although this will not surprise any teacher who uses the drills, it is an important step in a sober appraisal of the method. It demonstrates the effectiveness of the drills as such, independent of all the confounding factors of enthusiasm, personality, materials and measurement bias that are so difficult to control in classroom observations. It also transforms the basis of advocating such drills from a purely theoretical status of what "ought to be" considering what we know of the nature of language to the factual status of an experimental result.

The fact that our drilled subjects performed so much faster than the controls suggests that the method may be successful in developing relatively automatic habits and therefore of eliminating some of the stumbling, hesitating and puzzling so characteristic of beginners in foreign languages. The almost total absense of errors in word order in our Drill Group further suggests that automatic habit, rather than the conscious following of rules, is the "natural" basis of fluent speech. If grammatical speaking were a process of conscious rule-following, it would be expected that speed would be achieved at the expense of accuracy. Instead our faster group of subjects performed better.

The superiority of our drill subjects in vocabulary is not entirely explained. With number of presentations and responses to each word in training held constant for all subjects and with no evidence of differences in error rates for vocabulary during learning, we tentatively conclude that the difference in performance is caused by the presence of the additional task of putting the words into sentences, a task which was less practiced by the control than by the drill subjects. Another possible explanation lies in the difference between the learning and the testing conditions. The fact that the experimental subjects did significantly less well than the drills would suggest such an explanation. Some differences in the ease of learning in the two contexts may also exist, but we have no clear evidence of it in our data. For the language teacher it would be safe to conclude that vocabulary learned under pattern drill conditions is at least as well learned as in word lists and easier to use in later sentences.

It is also important for the language teacher to recognize what is not implied by data such as ours. In the first place,our conclusions about pattern drills apply only insofar as the drills resemble our experimental procedures. Pattern drills can take quite a variety of forms, and our results have no implications whatever about which would be best. We chose the form which we could adapt most easily to our experimental conditions. It seems likely that many other forms of drills would also be effective, perhaps more so than the one we chose.

Another caution for the teacher lies in the difference between the amounts and variety of knowledge learned in 15 hours in a lab and those that must be learned even in a single semester of a language course. For example, nothing is said or implied here about the more fundamental grammatical relationships suggested by transformational grammars. We have dealt only with a very superficial form of surface structure and only with very simple sentences. A great many linguistic skills are not even touched upon nor are their interactions with the processes we have studied. For example, it might be that excessive drilling of certain patterns would produce negative transfer effects on the learning of other patterns. In short, this experiment would not justify any blanket policy regarding all uses of pattern drills. It simply shows that under certain conditions, certain types of drills do produce the desired and expected results.

Conclusions from the viewpoint of the psychologist: The most important implication of our results for psychology is that sentence-forming skills have been produced which are not easily attributable to sequential responses. For psychology this is a radical conclusion. Almost all psychological theories of learning that have been applied to verbal and language learning depend fundamentally upon the assumption that sequential responses underly all language skills. Therefore before deciding finally that the traditional approach really does not explain this form of grammar learning, we should examine the possible interpretations of our results that might make them predictable from stimulus-response association theories.

The first would be some version of the "mediation" approach. Mediation, in this sense, is an attempt to account in terms of learned sequences for the development of associative responses that seem not to have been directly practiced. The basic idea of mediation is that once the stimulus-response sequences A - B and B - C have been learned, A will have developed a tendency to evoke the response C because the learned response B will function as a stimulus to evoke C. Most mediational theories further suppose that when A - B is learned, a backward response tendency is also formed such that B tends to evoke A as a response. Thus many secondary or incidental associations are formed in the course of learning and could account for the fact that learned behavior sequences occur without having themselves been practiced. In our experiment we might say that each y (test) word was linked with each x word of the class that followed (or preceded) it in the sentence pattern and through these mediating chains with each other y word of the same class as well. In this way sequential associative responses might be said to mediate our observed behavior sequences, at least in the case of the drill subjects. Our experimental subjects must be supposed to have learned a more complex chain involving the English translations of all the words. Although it has been amply demonstrated that both backward and mediated associative responses are possible, they are in many circumstances weaker than direct forward ones. Therefore we would expect our subjects to need a great deal of practice before the necessary connections were established. Our drill subjects were quite well practiced but they were so fast and accurate that it is difficult to believe that their performance in the test was due to such a complex network of separate sequential responses. However, we checked the mediation hypothesis by a comparison of the pattern of word association revealed in the free recall test. There, according to an association theory, order of recall should be determined by strength of associative connection. Table 30 compares for different groups of subjects the number of instances where a word was followed by another word of the same form class. For the most part this simply means the number of "repetitions" as in the previous analysis of clustering. These are approximately the "paradigmatic associations" of Ervin (1961). The second row contains the number of Ervin's "syntagmatic" associates, that is, words which are followed in free recall by words which could either precede or follow them in the sentence patterns learned. We included both forward and backward associations because both would be necessary for an associative mediation between two words of the same form-class in our patterns.

Mediation theory implies that mediated associative responses must be at best no stronger than the immediate associations that make them up. Since our test responses would have to be "mediated" by backward as well as forward associative responses, they should be even weaker than the forward responses involved. It is difficult by such a theory to account for the fact that our drill subjects, who had the largest number of experiences with the supposedly mediating connections, still show less than a third as many syntagmatic as paradigmatic associations in their recall sequences. This occurs in spite of the fact that we counted both forward and backward sequences as syntagmatic and there were therefore many more possible words that would be classified as syntagmatic responses.

Table 30

Mean Number of Sequences of Words in Free Recall Showing Grammatical Pattern

Group	Drill	Experimental	Control
Number of Repetitions (Words of same form-class following one another) (Paradigmatic associates)	18.12	13.81	15.19
Syntagmatic Associates (Forward and backward combined)	5.62	6.69	4.12

The main basis for mediation theory is probably theoretical rather than empirical. Sequential stimulus-response association has been assumed to be the basic building block of verbal learning since before the days of Ebbinghaus. Mediation is a modification made necessary by such empirical findings as ours. It has been shown by Kjeldergaard and Horton, (1961), among others, that such mediated associations are possible, but this does not by any means prove that they are in fact the basis of such things as paradigmatic associations. The data of Table 30, together with the complexity of the theoretical network of forward and backward associative responses that would be required in our case, suggest that it is time to open our minds to the possibility that forms of learning other than sequential associative responses are occurring in language learning.

Another attempt to make grammar learning like ours consistent with an association theory is Braine's idea of "contextual generalization". Here the mediating link is the context or sentence location. A closer look at what is meant here by "context" will show, however, that it cannot be specific mediating items or words. In Braine's experiment the new or generalized context into which nonsense words were correctly placed after learning had nothing specific in common with the previous context, that is, the context was an entirely new word, never before encountered. (The present experiment resembles Braine's in that no sequences were carried over from training into testing.) Braine points out that the only basis of "generalization" is the location in the sentence. This means we are dealing with an entirely different kind of generalization from that which implies either similarity or associative linkage of particular items.

In short, these two possible "sequential" explanations of our results are weak. Contextual generalization does not really bring any associative connections into the picture. Mediation would require so many and such complex sets of sequential associations that it is extremely difficult to believe that our few hours of training could produce the necessary selective reinforcements. Also, our free recall data show no evidence of these mediating sequential associations. In the absence of any prior assumptions to the effect that sequential associations must necessarily be found in order to account for any form of verbal learning, it would seem more plausible to interpret the results as showing the possibility of learning a more abstract pattern, a sequence of word classes, rather than of individual words. To accept such an explanation is to forego the possibility of talking of sequences of responses, however generalized. A form-class of words is not based upon any similarity of responses, that is, words with common privileges of occurrence do not necessarily have any physical resemblance to one another. Nor is it possible to assume that their associated "meanings" have any necessary resemblance. (The class of nouns, for example, goes much beyond the traditional names of persons, places or things.) A psychological interpretation of grammar learning seems to require that we abandon the simple associationism that has been assumed (though not actually tested or demonstrated) in much of the so-called "verbal learning" research. The present experiment may be added to the list of those that suggest the inadequacy of the assumption.

Although much theorizing in the field of verbal learning seems to ignore anything other than sequential association as a basic explanatory concept, other interpretations of learning have always been available. To mention only two recent instances, Gagné, in discussing the learning of concepts and principles, and Miller, in connection with the learning of grammar, both conclude that sequences of responses are not adequate basic elements of the learning in question.

Gagné (1965) defines "principles" as chains of concepts but points out that this is not the same as saying they are chains of stimulus and response connections. This is because concepts are not limited to particular stimuli nor to particular physical resemblances between stimuli which could form a basis of generalization.

Miller (1965) also proposes an entirely different approach to verbal learning. First he points out that to learn a language through reinforced sequences of responses would require much more than a lifetime because of the number of such particular sequences that are possible within the confines of one language. Second, he infers from the hierarchical nature of grammatical rules that, in principle, chains of successive responses cannot give any adequate account of the rules as they are now understood.

If we dismiss sequential association as an exclusive basis for a learning theory, we should next try to suggest a substitute model that will account for the learning we find. Miller makes several suggestions as to the kind of theory required (1965):

> "If we accept a realistic statement of the problem, I believe we will also be forced to accept a more cognitive approach to it: to talk about hypothesis testing instead of discrimination learning, about the evaluation of hypotheses instead of the reinforcement of responses, about rules instead of habits, about productivity instead of generalization, about innate and universal human capacities instead of special methods of teaching vocal responses, about symbols instead of conditioned stimuli, about sentences instead of words or vocal noises, about linguistic structure instead of chains of responses -- in short, about language instead of learning theory (p.20)."

Braine's contextual generalization has already been mentioned as a non-sequential model that might account for our findings. It is, however, limited in its application to the view of grammar as substitution. Although Braine has made some attempt to extend it to cover phrase structure, its usefulness there would depend upon assumptions about the further nature of phrase structure. It has not been thoroughly spelled out in this direction.

Some theory of learning through rules or principles seems to be required to account for behavior that conforms so exactly to rules. Gagné's (1965) notion of principles as hierarchies of concepts will serve as an example of such a theory. He defines concepts in a way that would appear to include such categories as form-classes in which there are not stimulus dimensions of similarity across the membership but which nevertheless have common properties, in this case grammatical properties. Formation of sentences could involve the application of a rule, a grammatical rule governing permissible sequences or structures.

One further requirement for a psychological theory of grammar learning remains, however. This is to account for the difference shown in this experiment between conscious application of rules, such as may develop through the vocabulary and grammar methods of learning, and the automatic, unconscious conformity to the rule that results from the pattern drill (or from the normal language learning process of children). Although the rule in our experiment was a simple one, there was nevertheless a considerable difference in the performance of our three groups. This led us to conclude that there was some important difference in the nature of their learning. Nevertheless all were able to perform in accordance with a rule. Our new theory, then, will have to make room for a distinction between the conscious application of a rule whose meaning is understood and the automatic, unconscious conformity to the same rule. The need for a theory of "unconscious" application of rules is also clear from the common observation that people, even pre-school children, speak grammatically though few if any know in any intellectual or conscious sense many of the rules they are following. The distinction between conscious and unconscious application of rules cannot be thrown under a behavioristic rug by saying that a concept or a rule is said to be "known" when it is correctly applied. The difference in behavior between our Control Group and the other groups requires a theory that will provide two quite different meanings for "applying a rule".

We must also provide in theory for the development of the ability to apply a rule unconsciously even when it may have started as a conscious process. Second-language learning is not the only instance where behavior which may begin by being directed consciously becomes quite unconscious in later application. Motor skills like driving a car and playing the piano could also be described as conscious and then unconscious application of complex sets of rules. Like language they are not fully accounted for by rigid sequences of motions, each triggering the next.

Whether we are dealing with rule-directed behavior that develops through originally conscious and intentional following of rules or with similar behavior for which there is no history of conscious rule-following, it must be recognized that a new concept of the organization of behavior is needed. Provision must be made in theory for the functioning of conceptual categories that probably are hierarchically organized and that are not identifiable with any particular responses or any particular stimulus characteristics. We need such a set of concepts not only to explain the grammatical patterning of language behavior, but probably also to understand its semantic functioning.

References

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 1953, 49, 229-240.

Braine, M. D. S. On learning the grammatical order of words. *Psychological Review*, 1963, 70, 323-348.

Brooks, N. *Language and language learning*. New York: Harcourt Brace & World, 1964.

Brown, R. & Berko, J. Word association and the acquisition of grammar. *Child Development*, 1960, 31, 1-14.

Carroll, J. B. & Sapon, S. M. *Modern language aptitude test manual*. New York: The Psychological Corporation, 1959.

Delattre, P. A technique of aural-oral approach. *French Review*, 1947, 20, 238-250.

Ebenholtz, S. M. Serial learning: position learning and sequential associations. *Journal of Experimental Psychology*, 1963, 66, 353-362.

Epstein, W. The influence of syntactical structure on learning. *American Journal of Psychology*, 1961, 74, 80-85.

Ervin, S. M. Changes in age in the verbal determinants of word-association. *American Journal of Psychology*, 1961, 74, 361-372.

Fraser, C., Bellugi, U. & Brown, R. Control of grammar in imitation, comprehension, and production. *Journal of Verbal Learning and Verbal Behavior*, 1963, 2, 121-135.

Fries, C. C. *Teaching and learning English as a foreign language*. Ann Arbor: University of Michigan Press, 1945.

Gagné, R. M. *The Conditions of Learning*. New York: Holt, Rinehart and Winston, 1965.

Kjeldergaard, P. M. & Horton, D. L. An experimental analysis of associative factors in mediated generalizations. *Psychological Monographs*, 1961, 75 (11, Whole No. 515).

Marks, L. E. & Miller, G. A. The role of semantic and syntactic constraints in the memorization of English sentences. *Journal of Verbal Learning and Verbal Behavior*, 1964, 3, 1-5.

Miller, G. A. Some preliminaries to psycholinguistics. *American Psychologist*, 1965, 20, 15-20.

Miller, G. A. & Isard, S. Some perceptual consequences of linguistic rules. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 217-228.

Miller, G. A. & Selfridge, J. A. Verbal context and the recall of meaningful material. American Journal of Psychology, 1950, 63, 176-185.

Palmer, H. E. The scientific study and teaching of languages. New York: World Book Co., 1917.

Politzer, R. L. Teaching French: an introduction to applied linguistics. Boston: Ginn and Company, 1960.

Politzer, R. L. Transfer of training. International Review of Applied Linguistics in Language Teaching, 1965, 3, 171-177.

Taylor, W. Cloze procedure: a new tool for measuring readability. Journalism Quarterly, 1953, 30, 415-433.

Appendix A

Experimental Materials: Six Russian Sentence Patterns

x	У ДЯ́ДИ	НЕТ	КНИ́Г
	by uncle	there are no	of books
	My uncle has no books.		
x	У ГОСПОЖИ́	НЕ́ БЫЛО	ПТИ́Ц
	by lady	was not	of birds
	The lady had no birds.		
x	ТАМ	МА́ЛО	ДЕТЕ́Й
	over there	few	of children
	There are few children over there.		
y	ЗДЕСЬ	МНО́ГО	КОРО́В
	here	many	of cows
	There are many cows here.		
y	У СОСЕ́ДКИ	НЕ БУ́ДЕТ	СЕСТЁР
	by neighbor	won't be	of sisters
	The neighbor won't have any sisters.		
y	У ДРУ́ГА	НЕМНО́ГО	СТУ́ЛЬЕВ
	by friend	some	of chairs
	The friend has some chairs.		

3-word Unlike-English Sentence Pattern

Pattern A I

Fig. 7

x	КУРИ́ТЬ	ВАМ	НЕЛЬЗЯ́
	to smoke	for you	not allowed
	You are not allowed to smoke.		
x	СЧИТА́ТЬ	МНЕ	МО́ЖНО
	to count	for me	possible
	It is possible for me to count.		
x	ПРА́ВИТЬ	ЕЙ	ТРУ́ДНО
	to drive	for her	difficult
	It is difficult for her to drive.		
y	ХОДИ́ТЬ	ИМ	НЕВОЗМО́ЖНО
	to walk	for them	impossible
	It is impossible for them to walk.		
y	УБЕЖА́ТЬ	ЕМУ́	ЛЕГКО́
	to run away	for him	easy
	It is easy for him to run away.		
y	СПАТЬ	НАМ	НУ́ЖНО
	to sleep	for us	necessary
	It is necessary for us to sleep.		

3-word Unlike-English Sentence Pattern

Pattern B II

Fig. 8

x	И́НОК	НА́ЧАЛ	ДЛИ́ННЫЙ	УЧЕ́БНИК
	monk	began	long	textbook

The monk began the long textbook.

x	ПИСА́ТЕЛЬ	ПО́НЯЛ	СКУ́ЧНЫЙ	ДОКЛА́Д
	writer	understood	boring	report

The writer understood the boring report.

x	ВРАЧ	ПИСА́Л	НЕМЕ́ЦКИЙ	О́ЧЕРК
	doctor	wrote	German	essay

The doctor wrote the German essay.

y	ОТЕ́Ц	ПЕРЕВЁЛ	КРА́ТКИЙ	РАССКА́З
	father	translated	short	story

Father translated the short story.

y	ЧЕЛОВЕ́К	ЧИТА́Л	СТА́РЫЙ	ДНЕВНИ́К
	person	read	old	diary

The person read the old diary.

y	УЧИ́ТЕЛЬ	КО́НЧИЛ	ГРУЗИ́НСКИЙ	УРО́К
	teacher	finished	Georgian	lesson

The teacher finished the Georgian lesson.

4-word Like-English Sentence Pattern

Pattern A III

Fig. 9

x	ЛЕ́ТОМ	ЖЕНА́	УХО́ДИТ	РА́НО
	in summer	wife	departs	early

In summer my wife departs early.

x	ОБЫ́ЧНО	УЧЁНЫЙ	МО́ЕТСЯ	ВЕ́СЕЛО
	usually	scholar	washes up	cheerfully

Usually the scholar washes up cheerfully.

x	О́СЕНЬЮ	ЧИНО́ВНИК	Е́ЗДИТ	МЕ́ДЛЕННО
	in autumn	official	rides	slowly

In autumn the official rides slowly.

y	ИНОГДА́	МОРЯ́К	ВСТАЁТ	НЕ́ХОТЯ
	sometimes	sailor	gets up	unwillingly

Sometimes the sailor gets up unwillingly.

y	ЗИМО́Й	СЛУЖА́НКА	ОДЕВА́ЕТСЯ	БЫ́СТРО
	in winter	maid	dresses	quickly

In winter the maid dresses quickly.

y	ВЕСНО́Й	ДВО́РНИК	ЕСТ	ПО́ЗДНО
	in spring	janitor	eats	late

In spring the janitor eats late.

4-word Like-English Sentence Pattern

Pattern B IV

Fig. 10

x	ВЫ́СТАВКУ	ОТКРЫВА́ЕТ	МНО́ГО	ХУДО́ЖНИКОВ
	exhibition	is opening	many	of artists

Many artists are opening the exhibition.

x	КОЛХО́З	ПОСЕЩА́ЕТ	МА́ЛО	ИНОСТРА́НЦЕВ
	collective farm	is visiting	few	of foreigners

Few foreigners are visiting the collective farm.

x	ПА́МЯТНИК	ЧИ́СТИТ	СО́РОК	ЛЮДЕ́Й
	monument	is cleaning	forty	of people

Forty people are cleaning the monument.

y	САД	ОСМА́ТРИВАЕТ	ПЯТЬ	ТОВА́РИЩЕЙ
	garden	is looking at	five	of comrades

Five comrades are looking at the garden.

y	СТОЛИ́ЦУ	И́ЩЕТ	НЕМНО́ГО	ДЕВУ́ШЕК
	capital	is searching for	some	of girls

Some girls are searching for the capital.

y	ЛА́ГЕРЬ	УВЕЛИ́ЧИВАЕТ	СТО	МА́ЛЬЧИКОВ
	camp	is enlarging	a hundred	of boys

A hundred boys are enlarging the camp.

4-word Unlike-English Sentence Pattern

Pattern B V

Fig. 11

x	ПРОВОДНИ́К	ЕГО́	НИГДЕ́	НЕ ВИ́ДЕЛ
	guide	him	nowhere	did not see
	The guide did not see him anywhere.			
x	ХОЗЯ́ИН	МЕНЯ́	НАРО́ЧНО	НЕ СЛУ́ШАЛ
	landlord	me	purposely	did not listen to
	The landlord purposely did not listen to me.			
x	ИВА́Н	ВАС	НИ РА́ЗУ	НЕ ИСКА́Л
	John	you	not once	did not look for
	John did not once look for you.			
y	РАБО́ТНИК	ИХ	ХОРОШО́	НЕ ОПИ́СЫВАЛ
	worker	them	well	did not describe
	The worker did not describe them well.			
y	СУДЬЯ́	НАС	НИКОГДА́	НЕ УЧИ́Л
	judge	us	never	did not teach
	The judge never taught us.			
y	МУЖ	ЕЁ	РА́НЬШЕ	НЕ ХВАЛИ́Л
	husband	her	earlier	did not priase
	The husband did not praise her earlier.			

4-word Unlike-English Sentence Pattern

Pattern A VI

Fig. 12